A SWITCHING REGRESSIONS FRAMEWORK FOR MODELS WITH COUNT-VALUED OMNI-DISPERSED OUTCOMES: SPECIFICATION, ESTIMATION AND CAUSAL INFERENCE

Wondimu Samuel Manalew

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Doctor of Philosophy
in the Department of Economics,
Indiana University

February 2020

Accepted by the Graduate Faculty of Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Doctoral Committee

Joseph V. Terza, Ph.D., Chair

Ben Boukai, Ph.D.

November 1, 2019

Una Osili, Ph.D.

Vidhura Tennekoon, Ph.D.

Matt Trombley, Ph.D.

ACKNOWLEDGEMENT

This dissertation has been done under close supervision of my advisor and dissertation committee chair, Dr. Joseph Terza, to whom I am extremely grateful for his endless support, advice and encouragement. Having him as an advisor was the best thing happened to me as a Ph.D. student. He often went out of his way to help me and he treated me with extreme patience. He has shown me, by his example, what a good researcher should be. Both as my professor and advisor he has trained me more than I could ever give him credit for here. Without his guidance, mentorship and persistent help, this dissertation would not have been possible.

I am indebted to the rest of my dissertation committee: Dr. Ben Boukai, Dr. Una Osili, Dr. Vidhura Tennekoon, and Dr. Matt Trombley for their helpful insights that improved the dissertation. I am especially grateful to two committee members with whom I have had the pleasure to co-author. I would first like to express my deepest appreciation to Dr. Tennekoon who has always been there for me whenever I needed help and advice. He is one of the most approachable, friendly, and helpful people I have ever met. Dr. Tennekoon is one of the key people behind my success as a Ph.D. student. I would like to extend my deepest appreciation to Dr. Osili from whom I have learned a lot about research, diligence and kindness.

I am also grateful to other faculty and staff members in the Department of Economics. My special appreciation goes to Dr. Steven Russell, Professor and Chair, for his tremendous support during my time in the department. My sincere thanks also go to Dr Ann Royalty, Dr. Wendy Morrison, Dr. Mark Ottoni-Wilhelm, Dr. Henry Mak, Dr. Jaesoo Kim, Dr. Sumedha Gupta, Dr. Subir Chakrabarti, for their wonderful instruction. I am also thankful to Terri Crews and Dana Ward for their help with various administrative affairs.

Finally, I would like to thank my beautiful, loving and supportive wife, Lidiya Woldeyes, and my precious child, Nehemiah, who provide unending inspiration. I would also like to thank my family members who own parents and siblings whose love and prayer are with me throughout this long journey.

Wondimu Samuel Manalew

A SWITCHING REGRESSIONS FRAMEWORK FOR MODELS WITH COUNT-VALUED OMNI-DISPERSED OUTCOMES: SPECIFICATION, ESTIMATION AND CAUSAL INFERENCE

In this dissertation, I develop a regression-based approach to the specification and estimation of the effect of a presumed causal variable on a count-valued outcome of interest. Statistics for relevant causal inference are also derived. As an illustration and as a basis for comparing alternative parametric specifications with respect to ease of implementation, computational efficiency and statistical performance, the proposed models and estimation methods are used to analyze household fertility decisions. I estimate the effect of a counterfactually imposed additional year of wife's education on actual family size (AFS) and desired family size (DFS) [count-valued variables]. In order to ensure the causal interpretability of the effect parameter as I define it, the underlying regression model is cast in a potential outcomes (PO) framework. The specification of the relevant data generating process (DGP) is also derived. The regression-based approach developed in the dissertation, in addition to taking explicit account of the fact that the outcome of interest is count-valued, is designed to account for potential sample selection bias due to a particular data deficiency in the count data context and to accommodate the possibility that some structural aspects of the model may vary with the value of a binary *switching variable*. Moreover, my approach loosens the equi-dispersion constraint [conditional mean (CM) equals conditional variance (CV)] that plagues conventional (poisson) count-outcome regression models. This is a particularly important feature of my model and method because in most contexts in empirical economics the data are either over-dispersed (CM < CV) or under-dispersed (CM > CV) – fertility models are usually characterized by the latter. Alternative count data models were discussed and compared using simulated and real data. The simulation results and estimation results using real data suggest that the estimated effects from my proposed models (models that loosen the equi-dispersion constraint, account for the sample selection, and accommodate variability in structural aspect of the models due to a switching variable) substantively differ from estimates from a conventional linear and count regression specifications.

Joseph V. Terza, Ph.D., Chair

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1: Introduction

The objective of most empirical economic research is to provide scientific evidence that can be used for policy purposes. Essential to this goal is rigorous specification and accurate estimation of parameters that characterize the key causal relationships. In this dissertation, I develop regression-based approaches to the specification and estimation of the effect of a presumed causal variable on a count-valued outcome of interest. I also derive statistics for relevant causal inference. The relevant treatment effect specification and estimation methods are all developed in a potential outcomes (PO) framework so as to ensure the causal interpretability of the targeted effect parameter and its estimate. Based on that PO (structural) model, I show, under certain conditions, that estimation of the relevant effect parameter can be based on the relevant data generating process (DGP). In addition to the PO specification, I also discuss the development of consistent estimates of the "deep" parameters of the relevant model, which are required for implementation of the policy effect estimator.

In this dissertation I use parametric count data regression models for estimating the policy effect of interest, and the models are cast in the PO framework. The PO specification in this dissertation is designed to 1) take explicit account of the fact that the outcome of interest is count-valued, unlike conventional linear regression models; 2) loosen the equi-dispersion constraint [conditional mean (CM) equals conditional variance (CV)] that plagues conventional (poisson) count-outcome regression models. This is a particularly important feature of my models and methods because in most contexts in empirical economics the data are either over-dispersed (CM < CV) or under-dispersed (CM > CV). Fertility data, which I use for illustration in this dissertation are usually characterized by under-dispersion; 3) accommodate the possibility that some structural aspects of models of *actual family size* (AFS) may vary with the value of a binary *switching variable*. For instance, in the survey data I use for illustration, some women, when questioned about *desired family size (DFS)*, did not give a numerical response. A substantial proportion of them (nearly 14%) responded with "At God's Will" (AGW). The PO framework proposed in this dissertation takes account of possible binary variation in regression structure that corresponds to differences in underlying attitudes toward fertility decision making that are reflected by responses to this survey question (AGW or not); 4) take account of a sample selection bias that arises in the context of estimating the effects of presumed causal variables (e.g. education) on DFS. For instance, women who give the AGW response to the desired fertility question are usually excluded from household desired fertility analysis. This produces a classic sample selection problem in modeling and estimation in the DFS context if such women systematically differ from the women who give numeric responses. The models and methods in this dissertation are designed to accommodate sample selection problems of this sort.

The remainder of the dissertation is organized as follows. In Chapter 2, I will: 1) give rigorous definition to the term "policy effect" and the ultimate estimation objective (the average incremental effect-AIE) ; 2) show how policy effects can be estimated in a regression framework; and 3) detail the conditions under which such policy effect parameters and their estimators can be interpreted as causal. As a special case, a causally interpretable specification for the effect of education on observed fertility will be developed along with a consistent estimator of that effect. Asymptotic inference will be discussed – both in the general case and in the context of estimating the effect of education on fertility. For estimating the AIE, flexible alternatives to the standard poisson specification, henceforth referred to as dispersion-flexible count-valued regression (DCR) models, will be considered. Therein, alternative conditional potential outcomes models (CPOMs) will be discussed and consistent AIE estimator will be derived for each DCR model. Finally, alternative DCR models will be compared in terms empirical measures of bias and efficiency to determine the DCR model that best accommodates under-dispersion. Both simulated and real data will be used for estimation and all comparisons will be in the context of estimating the AIE of an additional 1-year of education on AFS. In Chapter 3, a new count-data modeling and estimation framework will be developed to correct for the sample selection problem discussed earlier. This chapter will build on Chapter 2 in that it will incorporate the poisson generalizations (the DCR CPOMs) detailed therein. In Chapter 4, I develop a switching regression framework for the count data case to allow for a particular methodological concern that arises from an endogenous binary switching in the regression structure due to a response to a survey question that reflects fundamental differences in attitudes among women regarding fertility. In particular, I develop a causally interpretable switching regressions model of actual family size, which I will use to estimate the AIE of an additional year of education on observed fertility. Chapter 5 will summarize and conclude the discussion.

Chapter 2: Causal Inference Based on Dispersion-Flexible Count-Valued Regression (DCR)

In this chapter, the goal is to specify and estimate the causal effect of a presumed policy variable on a count-valued outcome using DCR models to accommodate under-dispersion. I begin with a general review of the PO framework as discussed in Terza (2019). After detailing relevant concepts within the PO framework, I will use the concepts to specify the AIE in count data context using the count data models, which I discuss in this chapter. As already noted, and as it will be clearer later, casting the specification and estimation of the AIE is important to make conditions required for causality explicit. I then present alternative DCR models by explicitly specifying their probability mass functions (pmf), followed by discussion of consistent estimation of the deep parameters from each regression model. I finally compare the alternative models based on AIE estimation using simulated and real data.

2.1 Specification of the Treatment Effect of Interest in the Potential Outcomes Framework

The main motivation for nearly all empirical economic research is to provide scientific evidence that can be used to assess causal relationships of interest based on relevant counterfactuals. Such assessments usually focus on the rigorous specification and accurate estimation of parameters that characterize the relationship between a *presumed causal variable of interest* ($\mathbf{X}$), whose value is to be exogenously set and altered in the context of the relevant counterfactual and a designated outcome of interest (

$\mathbf{Y}$).[1] Relationships of this type are typically characterized by an effect parameter (EP) and estimation of the EP is the objective of the empirical analysis. The main difficulty to be confronted in empirical estimation of an EP stems from the fact that it is *counterfactual* – i.e., it is based on random entities for which data are not directly observable. This is typically a substantial impediment to accurate estimation of the EP because the observable (*factual*) data are restrictive and, therefore, not directly representative of the relevant counterfactual entities. Terza (2019) details a general potential outcomes framework (GPOF) for EP specification, estimation and related inference that highlights the use of regression models and methods as means of reconciling the inherent disconnect between the counterfactual estimation objective and the factual data with which it is to be estimated.

First, I draw a distinction between the factual and counterfactual versions of the $\mathbf{X}$, and the $\mathbf{Y}$:

> X ≡ the random variable representing the observable (*factual*) version of the distribution of the $\mathbf{X}$ (The sampled values of the presumed causal variable are drawn from the distribution of X)

and

> X^* ≡ the random variable representing a hypothetical (*counterfactual*) exogenously mandated version of the distribution of the $\mathbf{X}$ (X^* is, by design, independent of all other variates germane to the present discussion).

[1]Henceforth, $\mathbf{X}$ and $\mathbf{Y}$ are to be taken as global replacements for the phrases "presumed causal variable of interest" and "outcome of interest," respectively.

Likewise, I distinguish two versions of the **Y**

> $Y \equiv$ the random variable representing the factual version of the distribution of the **Y** (The sampled values of the count-valued outcome are drawn from the distribution of Y)

and

> $Y_{X^*} \equiv$ the random variable representing the distribution of *potential outcome* that would have manifested for a particular X^* (an exogenously mandated version of the **X**).

Throughout the remainder of the discussion I will explicitly and implicitly reference a counterfactual in which the **X** is exogenously changed from X^{pre} to X^{post} (pre- to post-counterfactual). Without loss of generality I write $X^{post} = X^{pre} + \Delta$, where Δ is an observable random variable representing the counterfactual change in the distribution of the **X** for the relevant population. Note that X^{pre} and X^{post} are specific versions of X^*. Therefore, they are independent of all other variates germane to the discussion. So is Δ.

In the illustrative empirical analysis of household fertility decisions (discussed later) I seek to estimate the change in actual family size (the **Y**) that can be attributed to an additional year of the wife's education (the **X**). The components of the relevant counterfactual in this case are:

> $X^{pre} \equiv$ the random variable representing the pre-counterfactual distribution of education levels

and

$\Delta = 1 \equiv$ a one-year increment to the pre-counterfactual education level for each individual in the population,

Formally, I seek to estimate the following *average incremental effect* (AIE)

$$\mathrm{AIE}(\Delta) = \mathrm{E}[\mathrm{Y}_{\mathrm{X}^{\mathrm{pre}} + \Delta}] - \mathrm{E}[\mathrm{Y}_{\mathrm{X}^{\mathrm{pre}}}] \tag{1}$$

where $\Delta = 1$, $\mathrm{Y}_{\mathrm{X}^{\mathrm{pre}}} \equiv$ the PO corresponding to $\mathrm{X}^{\mathrm{pre}}$ and $\mathrm{Y}_{\mathrm{X}^{\mathrm{pre}}+\Delta} \equiv$ the PO corresponding to $\mathrm{X}^{\mathrm{pre}} + \Delta$.

2.2. Fully Parametric (FP) Specification of Models in the Potential Outcomes Framework

Effect parameters like in (1) are not directly estimable from data because $\mathrm{X}^{\mathrm{pre}}$ and $\mathrm{Y}_{\mathrm{X}^{\mathrm{pre}}}$ are counterfactual – they do not represent observable statistical populations from which samples can be drawn. As discussed by Terza (2019 and 2019b), this gap between the estimation objective (the inherently counterfactual EP) and the observable data (the data generating process (DGP) can be bridged via parametric specification of the conditional probability distribution of the relevant potential outcome $(\mathrm{Y}_{\mathrm{X}^*})$ given a vector of control covariates $(\mathrm{X}_{\mathrm{o}})$. To this end, following the approach proposed by Terza (2019), I posit the following FP structural specification for $(\mathrm{Y}_{\mathrm{X}^*} \mid \mathrm{X}_{\mathrm{o}})$

$$\mathrm{pmf}(\mathrm{Y}_{\mathrm{X}^*} \mid \mathrm{X}_{\mathrm{o}}) = \mathrm{f}_{(\mathrm{Y}_{\mathrm{X}^*} \mid \mathrm{X}_{\mathrm{o}})}(\mathrm{Y}_{\mathrm{X}^*}, \mathrm{X}^*, \mathrm{X}_{\mathrm{o}}; \pi) \tag{2}$$

where

$$\text{pmf}(Y_{X^*} \mid X_o) \equiv \text{the conditional pmf of } Y_{X^*} \text{ given } X_o$$

and the "deep" parameters of the model are π.

From (2) it follows that the conditional mean of Y_{X^*} is [2]

$$E[Y_{X^*} \mid X_o] = m(X^*, X_o; \pi) \tag{3}$$

where

$$m(X^*, X_o, \pi) = E[Y_{X^*} \mid X_o] = \int_{-\infty}^{\infty} Y_{X^*} f_{(Y_{X^*} \mid X_o)}(Y_{X^*}, X^*, X_o, \pi)\, dY_{X^*}$$

Using (3) and the law of iterated expectation (LIE), (1) can be written as

$$\text{AIE}(\Delta) = E[m(X^{pre} + \Delta, X_o; \pi)] - E[m(X^{pre}, X_o; \pi)] \tag{4}$$

It is relatively easy to show that under general conditions, given a consistent estimate of π (say $\hat{\pi}$) AIE can be consistently estimated using the following sample analog to (4)

$$\widehat{\text{AIE}(\Delta)} = \sum_{i=1}^{n} \frac{1}{n} \left\{ m(X_i^{pre} + \Delta_i, X_{oi}; \hat{\pi}) - m(X_i^{pre}, X_{oi}; \hat{\pi}) \right\} \tag{5}$$

where X_i^{pre} and Δ_i are the exogenously determined values of X^{pre} and Δ for the i^{th} observation in a sample of size n (i=1,..., n); and X_{oi} is the sampled value of X_o.

[2]See Terza (2019) for the derivation of the conditional mean function (3) from the joint pmf of $(Y_{X^*} \mid X_o)$ in (2).

2.3. Estimating the Treatment Effect of Interest

As is made clear by (5), consistent estimation of the EP in (4) hinges on the existence of a consistent estimate of π. With this in mind, and under Terza (2019) conditions [essentially, the conditional independence {CIND} of (Y_{X^*}) and X given X_o], the following is legitimate

$$\mathrm{pdf}(Y \mid X_o) = f_{(Y_{X^*} \mid X_o)}(Y, X, X_o; \pi) \tag{6}$$

In other words, under the CIND condition, the relevant DGP can be obtained from (2) by replacing the counterfactual random variables Y_{X^*} and X^* with the observable random variables Y and X, respectively. From (6) it follows that $\pi = [\beta_x \quad \beta_o^{'}]$ can be consistently estimated as the MLE obtained as

$$\hat{\pi} = \underset{\breve{\pi} = [\breve{\beta}_x \quad \breve{\beta}_o^{'}]}{\arg\max} \sum_{i=1}^{n} q(\breve{\pi}, Z_i) \tag{7}$$

where

$$q(\breve{\pi}, Z_i) = \ln[f_{(Y_{X^*} \mid X_o)}(Y_i, X_i, X_{oi}; \breve{\pi})] \tag{8}$$

and

$$Z_i = [Y_i \quad X_i \quad X_{oi}]$$

2.4. DCR-Based Causal Inference

As discussed before, the focus in this dissertation is on policy effect estimation for the case in which the outcome (the $\mathbf{Y}$) is count valued. Due to a number of shortcomings of conventional linear regression models (e.g. OLS) when used with count-valued outcome, in this dissertation, I use count-data models that take explicit account of the fact that $\mathbf{Y}$ is count-valued.[3] Equation (2) gives the generic count-outcomes PO specification; particular versions of this PO specification correspond with alternative forms for $\text{pmf}(Y_{X^*} \mid X_o)$. Below I discuss alternative forms for $\text{pmf}(Y_{X^*} \mid X_o)$.

2.4.1. The conventional Poisson specification

The most commonly used parametric count data model is the poisson. The relevant form of (2) for poisson case is given as

$$f_{(Y_{X^*} \mid X_o)}(Y_{X^*}, X^*, X_o, \pi) = \text{POI}(Y_{X^*}; \lambda^*) \tag{9}$$

where

$$\text{POI}(Y_{X^*}; \lambda^*) = \frac{(\lambda^*)^{Y_{X^*}} \exp(-\lambda^*)}{Y_{X^*}!}$$

[3] Conventional linear regression has a number of shortcomings when applied to count data. First, OLS does not account for the fact that count data are truncated at zero; thus, it could predict negative values for count-valued outcomes which are inherently restricted to positive (Wooldridge, 2010; Gardner et al., 1995). Second, since count data are skewed to the right, they are unlikely to satisfy the normality assumption of OLS, making statistical tests based on this assumption invalid (Cameron and Trivedi, 2005; Gardner et al., 1995). Third, the validity of hypothesis tests in the OLS also depends on assumptions about the homogeneity of variance of residuals that are unlikely to be met in count data (Winkelmann, 2008; Gardner et al., 1995; Cameron and Trivedi, 2005).

$$E[Y_{X^*} \mid X_o] = m(X^*, X_o, \pi) = \lambda^* = \exp(X^*\beta_x + X_o\beta_o) \tag{10}$$

and the parameter vector is $\pi = \beta$ with $\beta' = [\beta_x \quad \beta_o']$.

Using the LIE and combining (10) and (1), the relevant form of (4) for the poisson case is given as

$$\mathrm{AIE}(\Delta) = E\left[\exp[(X^{pre} + \Delta)\beta_x + X_o\beta_o)] - \exp(X^{pre}\beta_x + X_o\beta_o)\right] \tag{11}$$

As discussed in section 2.2, given consistent estimates of the deep parameters (say $\hat{\pi} = [\hat{\beta}_x \quad \hat{\beta}_o']$), (4) can be consistently estimated using the following sample analog

$$\widehat{\mathrm{AIE}(\Delta)} = \sum_{i=1}^{n} \frac{1}{n}\left\{\exp[(X_i^{pre} + \Delta_i)\hat{\beta}_{xi} + X_{oi}\hat{\beta}_o)] - \exp(X_i^{pre}\hat{\beta}_x + X_{oi}\hat{\beta}_o)\right\} \tag{12}$$

Turning now to consistent estimation of π, if the CIND condition holds, the relevant DGP as in (6) is

$$f_{(Y_{X^*} \mid X_o)}(Y, X, X_o, \pi) = \mathrm{POI}(Y; \lambda) \tag{13}$$

The needed consistent estimate of the deep parameters, $\hat{\pi} = [\hat{\beta}_x \quad \hat{\beta}_o']$, can now be obtained by the MLE based on this poisson version of the DGP in (6).

Although the poisson model accounts for the fact that the outcome is a non-negative integer, it is not without shortcomings. The *equidispersion* restriction under which the conditional mean of Y_{X^*} (and, therefore of Y) given X_o is equal to its conditional variance is imposed by the poisson pmf assumption. This assumption is often violated in real data applications (e.g. fertility data). The restrictiveness of the *equidispersion* condition has repercussions for the fertility literature as fertility data are frequently under-dispersed (see e.g., Winkelmann and Zimmermann, 1994; Wang and Famoye, 1997; Mayer and Riphahn, 2000; Cygan-Rehm, K, 2011; Islam et al. 2013; Mayer Riphahn, 2000). It has been shown that violation of this assumption will produce inconsistent estimates of the standard errors, rendering inference invalid (Gardner and Shaw, 1995). Winkelmann and Zimmermann (2009) show that while *over-dispersion* (conditional variance greater than the conditional mean) leads to a downward bias of the estimated standard errors, *under-dispersion (*conditional variance less than the mean) leads to an upward bias. [4]

More flexible count outcome models that account for all types of dispersion (DCR models) have been proposed in the literature. In the following section, I present the details of the most notable DCR models.

2.4.2. The Conway-Maxwell Poisson (CMP) Model

For CMP model, the relevant form of (2) is

$$f_{(Y_{X^*} \mid X_o)}(Y_{X^*}, X^*, X_o, \pi) = \mathcal{CMP}(Y_{X^*}; \lambda^*, \omega) \tag{14}$$

where

$$\mathcal{CMP}(Y_{X^*};\ \lambda^*, \omega) = \frac{(\lambda^*)^{Y_{X^*}}}{(Y_{X^*}!)^{\exp(\omega)}\, Z(\lambda^*, \exp(\omega))} \tag{15}$$

$$\lambda^* = \exp(X^*\beta_x + X_o\beta_o)$$

[4] A formal proof is given in Winkelmann and Zimmermann (1992).

$$Z(\lambda^*,\omega) = \sum_{j=0}^{\infty} \frac{\lambda^{*j}}{(j!)^{\exp(\omega)}}$$

The conditional mean of $(Y_{X^*} \mid X_o)$ for CMP model is

$$E[Y_{X^*} \mid X_o] = m(X^*, X_o, \pi) = \lambda^* \left(\sum_{j=1}^{\infty} \frac{j(\lambda^*)^{j-1}}{(j!)^{\exp(\omega)} Z(\lambda^*, \exp(\omega))} \right) \tag{16}$$

and the parameter vector is $\pi = [\beta \quad \omega]$ with $\beta' = [\beta_x \quad \beta_0']$ and $-\infty < \omega < \infty$.

Unlike the poisson, CMP has an additional parameter, ω -known as *dispersion parameter* -that measures variation in the data. The dispersion parameter allows the CMP model to account for under-dispersion and over-dispersion relaxing the equi-dispersion assumption of the poisson model. The data is over dispersed if $\omega < 0$, under-dispersed if $\omega > 0$, and equi-dispersed (not dispersed) if $\omega = 0$. Another nice property of the CMP is that it nests the standard poisson distribution when the dispersion parameter $\omega = 0$. The fact that CMP reduces to the poisson when $\omega = 0$ allows for a simple statistical test of under- or -over dispersion.

Using the LIE and combining (16) and (1) the relevant form of (4) for the CMP case is given as

$$AIE(\Delta) = E\left[m(X^{pre} + \Delta, X_o; \pi) - m(X^{pre}, X_o; \pi) \right] \tag{17}$$

Given consistent estimates of the deep parameters (say $\hat{\pi}$), (4) can be consistently estimated using the following sample analog

$$\widehat{\text{AIE}(\Delta)} = \sum_{i=1}^{n} \frac{1}{n}\left\{\hat{m}(X_i^{pre} + \Delta_i, X_{oi}, \hat{\pi}) - \hat{m}(X_i^{pre}, X_{oi}, \hat{\pi})\right\} \tag{18}$$

Turning now to consistent estimation of π, and noting that if the CIND condition holds, the relevant DGP as in (6) can be written as

$$f_{(Y_{X^*} \mid X_o)}(Y, X, X_o, \pi) = \mathcal{CMP}(Y; \lambda, \omega) \tag{19}$$

The consistent estimate of the deep parameters, $\hat{\pi}$, can now be obtained by the MLE based on this CMP version of the DGP in (6).

2.4.3. The Restricted Generalized Poisson (RGP) Model (Famoye,1993)

This model was introduced by Famoye (1993) and its application has been illustrated in Wang and Famoye (1997) for household fertility data and by Cui, Kim, and Zhu (2006) for modeling quantitative trait loci (Harris, 2012).

The relevant form of (2) for RGP is given as

$$f_{(Y_{X^*} \mid X_o)}(Y_{X^*}, X^*, X_o, \pi) = \mathcal{RGP}(Y_{X^*}; \lambda^*, \omega) \tag{20}$$

where

$$\mathcal{RGP}(Y_{X^*};\ \lambda^*, \omega) = \left(\frac{\lambda^*}{1+\omega\lambda^*}\right)^{Y_{X^*}} \frac{(1+\omega Y_{X^*})^{Y_{X^*}-1}}{Y_{X^*}!} \exp\left(\frac{-\lambda^*(1+\omega Y_{X^*})}{1+\omega\lambda^*}\right) \tag{21}$$

$$E[Y_{X^*} \mid X_o] = m(X^*, X_o, \pi) = \lambda^* = \exp(X^*\beta_x + X_o\beta_o) \tag{22}$$

Similar to the CMP, this model is a generalization of the standard poisson model and it is dispersion flexible with $\omega > 0$ and $\omega < 0$ indicating over-dispersion and under-dispersion respectively. RGP reduces to poisson for $\omega = 0$, which makes hypothesis testing of under-and over-dispersion straightforward. One limitation of this model, however, is that it is restricted with respect to the level of under-dispersion that it can accommodate; i.e. it must be true that the dispersion parameter ω is greater than (min(-1/max(λ),-1/max(Y)). This restriction is set because for some values of $\omega < 0$, the pmf of RGP does not sum to one.

Using the LIE and combining (22) and (1), the relevant form of (4) for the CMP case can be given as

$$\mathrm{AIE}(\Delta) = \mathrm{E}\left[\exp[(\mathrm{X}^{\mathrm{pre}}+\Delta)\beta_{\mathrm{x}} + \mathrm{X}_{\mathrm{o}}\beta_{\mathrm{o}})] - \exp(\mathrm{X}^{\mathrm{pre}}\beta_{\mathrm{x}} + \mathrm{X}_{\mathrm{o}}\beta_{\mathrm{o}})\right] \tag{23}$$

Given consistent estimates of the deep parameters (say $\hat{\pi}$), (4) can be consistently estimated using the following sample analog

$$\widehat{\mathrm{AIE}(\Delta)} = \sum_{i=1}^{n} \frac{1}{n}\left\{\exp[(\mathrm{X}_{i}^{\mathrm{pre}}+\Delta_{i})\hat{\beta}_{\mathrm{x}i} + \mathrm{X}_{\mathrm{o}i}\hat{\beta}_{\mathrm{o}})] - \exp(\mathrm{X}_{i}^{\mathrm{pre}}\hat{\beta}_{\mathrm{x}} + \mathrm{X}_{\mathrm{o}i}\hat{\beta}_{\mathrm{o}})\right\} \tag{24}$$

Turning now to consistent estimation of π, with CIND conditions, the relevant DGP as in (6) can be written as

$$f_{(Y_{X^*} \mid X_o)}(Y, X, X_o, \pi) = \mathcal{RGP}(Y; \lambda, \omega) \tag{25}$$

The consistent estimate of the deep parameters, $\hat{\pi}$, can now be obtained by the MLE based on this RGP version of the DGP in (6).

2.4.4. The Hyper Poisson Model (HP)

The Hyper-Poisson is another DCR model that can handle both over- and under-dispersion. Like the other DCR models discussed above, HP is a generalization of the standard poisson model. In this case the generalization is made such that the regressors are introduced in the mean and at the same time influence the over-or under-dispersion behavior of the distribution (Sáez-Castillo and Conde-Sánchez, 2013). Using over-dispersed and under-dispersed motor vehicle crash data from Toronto and Korea, Khazraee (2014), shows this model fits well in both cases.

The relevant form of (2) for HP is given as

$$f_{(Y_{X^*} \mid X_o)}(Y_{X^*}, X^*, X_o, \pi) = \mathcal{HP}(Y_{X^*}; \lambda^*, \omega) \tag{26}$$

where

$$\mathcal{HP}(Y_{X^*};\ \lambda^*, \omega) = \frac{1}{K(1, \omega, \lambda^*)} \frac{\lambda^{*Y_{X^*}}}{(\omega)_{Y_{X^*}}} \tag{27}$$

$$= \frac{1}{K(1, \exp(\nu), \lambda^*)} \frac{\lambda^{*Y_{X^*}}}{(\exp(\nu))_{Y_{X^*}}}$$

where $\omega > 0$; $\nu = \ln(\omega)$; $\lambda^* = \exp(X^*\beta_x + X_o\beta_o)$

$$(a)_r = \frac{\Gamma(a+r)}{\Gamma(a)} \tag{28}$$

for a > 0 and r a positive integer, and

$$K(a,b,c) = \sum_{r=0}^{\infty} \frac{(a)_r c^r}{(b)_r r!}. \tag{29}$$

Like the other DCR models HP has an extra parameter, ω, that measures variability I the data. This distribution is over-dispersed if $\omega > 1$, is the poisson distribution if $\omega = 1$ and is under-dispersed if $\omega < 1$. The conditional mean of the outcome in this case is given as

$$E[Y_{X^*} \mid X_o] = m(X^*, X_o, \pi) = \lambda^* - (\exp(\nu) - 1)\left(\frac{K(1, \exp(\nu), \lambda^*) - 1}{K(1, \exp(\nu), \lambda^*)}\right) \tag{30}$$

Using the LIE and combining (30) and (1) the relevant form of (4) for the CMP case can be shown to be given as

$$AIE(\Delta) = E\left[m(X^{pre} + \Delta, X_o; \pi) - m(X^{pre}, X_o; \pi)\right] \tag{31}$$

Given consistent estimates of the deep parameters (say $\hat{\pi}$) consistent estimate of (4) can be obtained using the following sample analog

$$\widehat{AIE(\Delta)} = \sum_{i=1}^{n} \frac{1}{n}\left\{\exp[(X_i^{pre} + \Delta_i)\hat{\beta}_{xi} + X_{oi}\hat{\beta}_o)] - \exp(X_i^{pre}\hat{\beta}_x + X_{oi}\hat{\beta}_o)\right\} \tag{32}$$

Turning now to consistent estimation of π, if the CIND condition holds, the relevant DGP as in (6) can be written as

$$f_{(Y_{X^*} \mid X_o)}(Y, X, X_o, \pi) = \mathcal{HP}(Y; \lambda, \omega) \tag{33}$$

The consistent estimate of the deep parameters, $\hat{\pi}$, can now be obtained by the MLE based on this HP version of the DGP in (6).

2.5. Comparison of Alternative Dispersion Flexible Conditional Potential Outcomes Models (CPOMs) using Simulated Data

To measure the cost of ignoring under-dispersion and over-dispersion, I have generated data with various dispersion levels from RGP, CMP and HP models for the count outcome Y_{X^*}, with policy variable X^*, other regressors X_o, and a constant term. The data simulation was conducted according to the following sampling design[5]. I simulate data with sample size = 1, 000 and number of repetitions=1,000. Simulation of the covariates, $X=[X^* \ X_o \ 1]$ is designed such that X^* and X_o are uniformly distributed with both mean 1 and variance of 3. The choice of the true regression parameters is $\beta_1' = \beta_2' = [\beta \quad \beta_o \quad c] = [.25 \quad .25 \quad -.25]$ where, c denotes the true intercept. The dispersion parameter ω is specified such that all the models generate under-dispersed count data at various levels of dispersion.[6] The 'true' *average incremental effect* (AIE) is conducted by taking a super sample of 2,000,000 observations generated using the sample design and plugging the true parameters into (4) using simulated data with $m(\ )$ defined as in (3). Following the sample data generation and approximate calculation of the 'true' AIE, estimation of the effect of the policy variable X^* (*the average incremental effect*) was conducted using the poisson model (a model that

---

[5] Stata/Mata was used for the data simulation and estimation.

[6] To generate under-dispersed data using RGP, I choose $\omega < 0$. And $\omega < 1$ and $\omega > 1$ are used for HP and CMP cases, respectively.

ignores the under-dispersion in the data) and all the DCR models (models which take account of the under-dispersion) using

$$\widetilde{\mathrm{AIE}}_{m} = \sum_{r=1}^{R} \frac{1}{R} \widehat{\mathrm{AIE}}_{mr} \tag{34}$$

where m=Poisson, CMP, RGP or HP

For the estimation of the AIE of the policy variable, X^*, I set the exogenous policy change at Δ =1; e.g. one additional year of education using (5). These estimates of AIEs from each of the models are compared with the true value, denoted by AIE(Δ_L), which is calculated by substituting the true parameters specified in the sampling design into (3). Performances of alternative models are compared taking and Absolute Proportional Bias (APB) and Average Absolute Proportional Bias (AAPB) as measures of goodness-of-fit. These measures are defined in (35) and (36), respectively.

$$\widetilde{\mathrm{APB}}_{m} = \frac{\mathrm{abs}(\widehat{\mathrm{AIE}} - \mathrm{AIE})}{\mathrm{abs}(\mathrm{AIE})} \tag{35}$$

$$\widetilde{\mathrm{AAPB}}_{m} = \sum_{r=1}^{R} \frac{1}{R} \left(\frac{\mathrm{abs}(\widehat{\mathrm{AIE}}_{mr} - \mathrm{AIE})}{\mathrm{abs}(\mathrm{AIE})} \right) \tag{36}$$

2.5.1. Estimation Results

Table 1 presents the AIE estimates from the poisson and the alternative DCR models at different values of the dispersion parameter, using data generated from the CMP model. The absolute proportional bias from the poisson model is by far larger than the bias from the DCR models, showing the cost of ignoring the under-dispersion in the data. I also find that that the magnitude of the absolute proportional bias increases as the level of under-dispersion increases. For example, the AAPB bias reaches 37% when $\omega = 10$. Not surprisingly, the RGP model misbehaves for some values of the dispersion parameter and thus the estimation faces convergence problems. This is because of the restriction the model puts on the range of values the dispersion parameter takes. The other models produce estimates for all values of the dispersion parameter. Table 2 presents estimation results using the RGP data. Similar to the CMP data case, there is an increase in bias when poisson model is used. Table 3 gives results from HP model. The HP model is the best in terms of absolute percentage bias in this case. Poisson and CMP are performing worse. However, the differences in the bias the models produce are not very large.

2.6. Comparison of Alternative DCR CPOMs Using Real Data

In addition to simulated data, real data from the 1989 wave of the Michigan Panel Study of Income Dynamics (PSID) is used to make model comparisons. This data was first used by Wang and Famoye (1997) to illustrate their household fertility choice using the RGP model. The data includes married women aged between 18 and 40 who are not heads of households and with nonnegative total family income. The outcome variable, the total number of children up to 17 years old in a family, is a nonnegative integer ranging from zero to nine in the sample (see Table 4 for descriptive statistics). This application is well suited for my illustration for two reasons: first, the outcome is count valued; second, it has been shown by Wang and Famoye (1997) that this data exhibits under-dispersion. I have also confirmed that the data is under-dispersed. Employment status of the women, age, education, family income, race and area of residence are the chosen explanatory variables.

The estimation of the coefficient parameters and the policy effect (AIE) is then conducted by Poisson, RGP, COM-Poisson and Hyper-Poisson models. The summarized estimates of parameters are given in Table 5. Results from all the models show that the education level of the mother is statistically significant and is inversely related to the number of the children in household. The estimated dispersion parameters from all of the models indicate under-dispersion in the data.[7] And the asymptotic-t-statistics show the significance of the dispersion parameter.[8] Comparing the models considering the parameters estimates, Table 5 shows that the parameter estimates from the poisson regression and the DCR models are qualitatively similar. Here, ignoring the under-dispersion in the data is not causing bias. This result is not unexpected because of the quasi maximum likelihood estimator (QMLE) property of the poisson model, i.e. if the conditional mean is correctly specified, the poisson model produces consistent estimates even in the presence of under-dispersion.

---

[7] It should be noted here that for CMP and HP models the dispersion parameter obtained from Stata output is $\nu = \ln(\omega)$.

[8] The null hypothesis for all cases is Poisson. But Poisson is when $\omega = 0$ for RGP case and when $\nu = 0$ for CMP and HP cases. The alternative hypothesis is set accordingly.

However, the main interest is not in the parameter estimates as the estimates tell us nothing about the policy effects of the variables; they only indicate the significance of each explanatory variable and their contribution to the conditional mean. For the purpose of policy analysis, I estimate the AIE as defined in (4), using a consistent estimator defined in (5). These results are presented Table 6. Similar to the case of the parameter estimates, AIE estimates from the poisson model are close to the estimates of the DCR models. In terms of precision, however, the DCR models outperform the poisson model. If precision is measured using standard errors, for example, RGP, COM and HP yield nearly a 15% increase in precision relative to the poisson model. Moreover, the COM-poisson t-test is 17% higher than the one for the poisson estimate. When count data are under-dispersed, therefore, the estimates from poisson regression are consistent but there exist substantial differences in precision AIE estimates from poisson and the DCR models.

2.7. Summary, Discussion and Conclusion

Using both simulated data and real data I examined and compared alternative DCR models with respect to their estimation precision and bias when applied to under-dispersed data with count-valued outcome. Both the simulation results and the real data analysis show that, overall, the DCR models outperform the poisson model based on the precision criteria detailed earlier. This result may suggest that inference based on the poisson model could be misleading when there is significant dispersion in the data.

Chapter 3: Accounting for Sample Selection Bias in DCR

The econometric issue of *sample selection* concerns the possible biases that arise when a nonrandomly sampled set of observations from a population is used as if the sample were random to make inferences about that population (Greene, 2001). In this chapter, I am interested in selection issues in the context of count-valued outcome regression models. The main objective is to devise a general fully parametric sample selection modeling framework for DCR models discussed in Chapter 2. The sample selection framework developed incorporates the PO framework analogous to the framework developed in Chapter 2 to take account of the counterfactual nature of the effect parameter (EP) and it provides a means to clearly and coherently define the EP. The sample selection models were evaluated and compared against models that ignore selection issues using simulated data.

To illustrate the econometric issues, I considered an empirical analysis of the effect of a 1-year increment in wife's education on desired family size (DFS). The data for this analysis comes from the 2016 Ethiopia Demographic and Health Survey (EDHS). The data includes information collected from 35,081 women between 15 to 49 years of age. The data about the desired number of children were based on responses to the survey question: "*If you could go back to the time you did not have any children and could choose exactly the number of children to have in your whole life, how many would that be?*" The respondents had options to either give numeric or non-numeric responses to the survey question. While 86 percent gave numeric responses, the 14 percent gave the "At God's Will" (AGW) response.

The presence of the AGW response is viewed as a methodological problem. One of the most common treatments of such responses is to categorize all responses as an ordinal scale and to include nonnumeric responses in the highest preference group (Riley, 1993). Other strategies include treating nonnumeric responses as missing data and excluding them from the analysis (Riley, 1993; Jensen, 1985). Empirical evidences suggest, however, that treating these responses as missing data and dropping them from analyses will bias results (Jensen,1985). Jensen (1985) dealt with the selection bias issue for the case in which the outcome of interest is DFS using the Heckman (1976) sample selection for linear regression model. He argues that the empirical evidence is insufficient to show that the true preferences of nonnumeric responders are different from those of women who give numeric responses. Jensen's approach clearly, however, ignores the fact that the dependent variable—*DFS—is* count valued and that the data for the outcome variable could be over or under- dispersed. The PO framework presented in this chapter is designed to accommodating the sample selection problem simultaneously taking care of the under-dispersion and taking explicit account of the fact that the outcome is count-valued.

3.1. Count-Valued Regression-Based Potential Outcomes Modeling in the Presence of Endogenous Sample Selection

I follow the GPOF discussed in Chapter 2, to specify, identify and estimate parameters that characterize the causal relationship between the $\mathbf{X}$ and the $\mathbf{Y}$[9]. The EP of interest in this chapter, like in Chapter 2, is the AIE. Unlike the case in Chapter 2,

[9] See Chapter 2 for definitions of distinct versions of $\mathbf{X}$ and $\mathbf{Y}$.

however, the objective here is to estimate the AIE of a policy-driven change in education (Δ) on desired family size (DFS). The AIE here is defined as

$$\mathrm{AIE}_2(\Delta) = \mathrm{E}[\mathrm{Y}_{\mathrm{X}^{\mathrm{pre}} + \Delta}] - \mathrm{E}[\mathrm{Y}_{\mathrm{X}^{\mathrm{pre}}}]. \tag{37}$$

In the context of the illustrative example of fertility analysis

$\mathrm{X}^{\mathrm{pre}}$ ≡ the exogenously imposed pre-policy distribution of education levels

Δ ≡ the exogenously imposed distribution of policy-induced changes in education levels

$\mathrm{Y}_{\mathrm{X}^{\mathrm{pre}}}$ ≡ the potential DFS corresponding to $\mathrm{X}^{\mathrm{pre}}$

and

$\mathrm{Y}_{\mathrm{X}^{\mathrm{pre}}+\Delta}$ ≡ the potential DFS corresponding to $\mathrm{X}^{\mathrm{pre}} + \Delta$.

As discussed in Chapter 2, however, the EP in (37) is not directly estimable from data because the relevant versions of $\mathrm{Y}_{\mathrm{X}^*}$ (viz., $\mathrm{Y}_{\mathrm{X}^{\mathrm{pre}} + \Delta}$ and $\mathrm{Y}_{\mathrm{X}^{\mathrm{pre}}}$) in (37) are at least partially counterfactual and, therefore, not fully observable via sampling. Following the approach in Terza (2019 and 2019b), I use a FP specification of the conditional probability distribution of the relevant potential outcome $(\mathrm{Y}_{\mathrm{X}^*})$ given a vector of observable control covariates $(\mathrm{X}_{\mathrm{o}})$ and a vector of unobservable confounders $(\mathrm{X}_{\mathrm{u}})$ to establish the connection between the EP and the relevant data generating process (DGP).

For the FP structural specification of $(Y_{X^*} \mid X_o, X_u)$, I am interested in the case in which the systematic observability of the outcome (Y_{X^*}) depends on sample selection indicator defined as

$$S = I(W\delta + X_u > 0) \tag{38}$$

where S is the selection variable that determines observability or non-observability of the count-valued outcome; $I(A)$ denotes the indicator function that takes the value 1 if condition A is true and 0 otherwise. W is a vector of observable regressors, i.e. $W = [X \mid W^+]$ and W^+ is a vector (scalar) of *identifying instrumental variables (IV)*, δ is coefficient vector associated to the observable covariates, and X_u represents the unobservable factors that affect the probability of being selected. The distribution of $(X_u \mid W)$ is assumed to be known.

If the unobservable (and, therefore, uncontrollable) factors in the sample selection rule (X_u) in (38) are correlated with unobservable determinants of the outcome, Y_{X^*}, then causal (policy) effect estimates based on parameter estimates that ignore this correlation will be biased.

In the illustrative application in this chapter, for example, I am interested in estimating the AIE of a 1-year increment in wife's education on DFS. Suppose that X_u represents a measure of non-religiosity so that less religious women are more likely to be selected into the sample $S = 1$ – i.e. they are less likely to give an AGW response to survey response. If it is also true that for less religious women the effect of education on fertility is larger (in absolute value), then restricting the analysis to those for whom $S = 0$

(those who did not respond AGW) will likely overstate the effect of education on the fertility decisions (in absolute value) among females at large. This is an example of what is typically called *sample selection bias*.

Turning to the FP specification of $(Y_{X^*} \mid X_o, X_u)$, I note, again, that values of the outcome are not observable for all members of the relevant population and sampling preclusion is not random. Instead, the observability and non-observability of Y_{X^*} is governed by a systematic sample selection rule (38) which is determined by both observable and unobservable factors. It is convenient, in this case, to define Y^*, a partially qualitative variable, whose outcome is real and observable if $S = 1$ and qualitative ("not observed") if $S = 0$. In the context of the illustrative example of household fertility analysis, the partially qualitative variable Y^* takes count values 0, 1, 2, … if S=1, and the 'AGW' response if $S = 0$. The conditional pdf of the partially qualitative outcome variable Y^* given X, X_o, X_u and W^+ can be defined in a potential outcome framework as

$$h(Y^*_{X^*} \mid W, X_u, S; \pi, \delta) = \begin{cases} f_{(Y^*_{X^*} \mid X_o, X_u)}(Y_{X^*} \mid X^*, X_o, X_u; \pi) & \text{if} \quad S = 1 \\ 1 & \text{if} \quad S = 0 \end{cases} \tag{39}$$

Following the approach in Terza (2009), I base estimation of the parameters on the joint pmf of the partially qualitative variable Y^* and S conditional on W. It can be shown that the joint pmf is given by

$$h(Y^*_{X^*}, S \mid W;\, \pi,\, \delta) = \left(\int_{-W\delta}^{\infty} f(Y^*_{X^*} \mid W, X_u, S; \pi)\;\; g(X_u)\; dX_u \right)^S \times G(-W\delta)^{1-S} \quad (40)$$

Combining (38) and (40), it can be shown that the joint probability density function is

$$\begin{aligned} \mathrm{pmf}(Y^*_{X^*}, S \mid W) &= h_{(Y^*_{X^*}, S \mid W)}(Y^*_{X^*},\, S,\, W,\, \pi,\, \delta) \\ &= \left(\int_{-W\delta}^{\infty} f_{(Y^*_{X^*} \mid X, X_o, X_u)}(Y^*_{X^*}, X^*, X_o, X_u;\, \pi)\;\; g(X_u)\; dX_u \right)^S \\ &\quad \times G(-W\delta)^{1-S} \end{aligned} \quad (41)$$

where g( ) and G( ) denote the known pdf and cdf of X_u, respectively, the functional form of $f_{(Y^*_{X^*} \mid X_o, X_u)}(Y^*_{X^*},\, X^*,\, X_o, X_u;\, \pi)$ is known, π is the vector of "deep" parameters of the model, and X_u is an unobservable scalar regressor variate. [10]

Knowledge of the functional form of $f_{(Y^*_{X^*} \mid X_o, X_u)}(Y^*_{X^*},\, X^*,\, X_o, X_u;\, \pi)$ in (41) implies knowledge of the conditional mean of $Y^*_{X^*}$, viz.

$$E_2[Y^*_{X^*} \mid X_o,\, X_u] = m(X^*,\, X_o,\, X_u;\, \pi) \quad (42)$$

where the functional form of $m(\;)$ is known, X_u is not a confounder for X (in fact, X_u and $X = [X \quad X_o]$ are assumed to be independent), and pdf and cdf of X_u are known and

[10] Note that X_u is integrated out in (41).

are denoted by $g(X_u)$ and $G(X_u)$, respectively, with $E(X_u) = 0$.

Using (42) and the law of iterated expectation (LIE), (37) can be re-written as

$$\begin{aligned} AIE_2(\Delta) = E_{X^{pre}, X_o, \Delta}\Bigg[\int_{-\infty}^{\infty} \Big\{ & m(X^{pre} + \Delta, X_o, X_u; \pi) \\ & - m(X^{pre}, X_o, X_u; \pi) \Big\} g(X_u \mid W)\, dX_u \Bigg] \end{aligned} \tag{43}$$

It is relatively easy to show that under general conditions, given a consistent estimate of π (say $\hat{\pi}$) AIE can be consistently estimated using the following sample analog to (43)

$$\begin{aligned} \widehat{AIE_2(\Delta)} = \sum_{i=1}^{n} \frac{1}{n} \int_{-\infty}^{\infty} \Big\{ & m(X_i^{pre} + \Delta_i, X_{oi}, X_u; \hat{\pi}) \\ & - m(X_i^{pre}, X_{oi}, X_u; \hat{\pi}) \Big\} g(X_u \mid W) dX_u \end{aligned} \tag{44}$$

where X_i^{pre} and Δ_i are the exogenously determined values of X^{pre} and Δ for the i^{th} observation in a sample of size n (i=1,..., n); and X_{oi} is the sampled value of X_o.

3.2. Consistent Estimation of the Deep Parameters of the Model

It is clear from (44) that consistent estimation of the EP in (43) hinges on the existence of a consistent estimate of π. As discussed earlier, given the fully parametric specification, the estimation of the deep parameters is conducted by MLE. However, the FP specification in (41) is given in terms of potential outcome $(Y^*_{X^*})$ and exogenously mandated policy variable (X^*) which are both counterfactual. Under Terza (2019) conditions [essentially, CIND of $(Y^*_{X^*})$ and X given X_o and X_u], however, the counterfactuals $(Y^*_{X^*})$ and (X^*) can be replaced by the partially observable version Y^* and X, respectively in (41). Observable version of the FP specification (DGP) can, therefore, be written as

$$\mathrm{pmf}(Y^*, S \mid W) = h_{(Y^*_{X^*}, S \mid W)}(Y^*, S, W, \pi, \delta)$$

$$= \left(\int_{-W\delta}^{\infty} f_{(Y^*_{X^*} \mid X, X_o, X_u)}(Y^*, X, X_o, X_u; \pi)\; g(X_u)\; dX_u \right)^S$$

$$\times G(-W\delta)^{1-S} \qquad (45)$$

It follows now that π and δ can be consistently estimated as the MLE obtained as

$$\hat{\pi} = \underset{\breve{\pi} = [\breve{\beta}_X \;\; \breve{\beta}_o' \;\; \breve{W} \;\; \breve{\delta}]}{\arg\max} \sum_{i=1}^{n} q(\breve{\pi}, Z_i) \qquad (46)$$

where

$$q(\breve{\pi}, Z_i) = \ln[f_{(Y^*_{X^*} \mid X_o, X_u)}(Y_i, X_i, X_{oi}, X_{ui}; \breve{\pi})] \qquad (47)$$

and

$$Z_i = [Y_i, X_i, X_{oi}, X_{ui}]$$

Note that as there are non-closed-form integrals in the log-likelihood function (46), Gauss-Legendre quadrature will be used to numerically approximate them, which adds more computational complexity into the implementation of the estimator. It is important to acknowledge this because one criterion of assessing different models is their computational efficiency. Inference using these estimates can be conducted via standard asymptotic theory.

To correct for sample selection bias in the estimation of the deep parameters and thus the EP of interest, I specify that $W = [X_o \quad W^+]$, where W^+ is a vector of instrumental variables (IV) satisfying the following conditions

(1) $E[X_u \mid W] = 0$

(2) Exclusion restriction:

$$f_{(Y|X,X_o,X_u)}(Y, X, X_o, X_u; \pi) = pdf(Y \mid X, X_o, X_u) = pdf(Y \mid X, W, X_u)$$

which implies

$$m(X, X_o, X_u; \pi) = E[Y \mid X, X_o, X_u] = E[Y \mid X, W, X_u]$$

and (3) $COV(X, W)$ is sufficiently different from zero.

3.3. Alternative Dispersion Flexible Conditional Potential Outcomes Models (CPOMs) with Sample Selection

Functional forms used here are very similar to the ones in Chapter 2 except that the functional forms in this chapter are designed such that they allow for possible unobservable conditioning variates (X_u). All of the functional forms discussed in Chapter 2 can be implemented for the sample selection case just by redefining the mean related parameter, λ, to incorporate X_u as a regressor. I consider the poisson model and two CPOMs (CMP and HP models).

3.3.1. Poisson Model

The FP regression-like structural specification for the potential outcome ($Y^{*}_{X^{*}}$), which is the relevant form of (40) for poisson case, is given as:[11]

$$\mathrm{pdf}(Y^{*}_{X^{*}}, S \mid W) = \left(\int_{-W\delta}^{\infty} \mathrm{POI}\left(Y^{*}_{X^{*}}, \lambda^{*}\right) \; g(X_u)\, dX_u \right)^{S} \times G(-W\delta)^{1-S} \tag{48}$$

where

$$\mathrm{POI}\left(Y^{*}_{X^{*}}, \lambda^{*}\right) = \frac{\lambda^{Y^{*}_{X^{*}}} \exp(-\lambda^{*})}{Y^{*}_{X^{*}}!}$$

and

$$E[Y^{*}_{X^{*}} \mid X^{*}, X_o, X_u] = m(X^{*}_{p}, X_o, X_u;\, \pi) = \lambda = \exp(X^{*}\beta_x + X_o\beta_o + X_u\beta_u)\,. \tag{49}$$

Using the LIE and combining (49) and (37) the relevant form of (40) for the poisson case is

$$\begin{aligned} \mathrm{AIE}_2(\Delta) = E_{X^{pre}, X_o, \Delta}\Bigg[\int_{-\infty}^{\infty} \Big\{ & m(X^{pre} + \Delta,\, X_o,\, X_u;\, \pi) \\ & - m(X^{pre},\, X_o,\, X_u;\, \pi) \Big\}\, g(X_u \mid W)\, dX_u \Bigg] \end{aligned} \tag{50}$$

Given a consistent estimate of π (say $\hat{\pi}$), the AIE can be consistently estimated using the following sample analog to (50)

[11] See the functional form specification for poisson for no sample selection case in Chapter 3 for comparison.

$$\widehat{\mathrm{AIE}_2(\Delta)} = \sum_{i=1}^{n} \frac{1}{n}\Big[\mathrm{m}(X_i^{pre} + \Delta_i,\, X_{oi},\, X_u;\, \hat{\pi}) - \mathrm{m}(X_i^{pre},\, X_{oi},\, X_u;\, \hat{\pi})\Big\}\, g(X_u \mid W)\, dX_u \Big] \tag{51}$$

Turning now to consistent estimation of π, if the CIND condition holds, the relevant DGP as in (6) can be given as

$$f_{(Y_{X^*}^* \mid X_o)}(Y,\, X,\, X_o,\, \pi) = \mathrm{POI}(Y;\, \lambda) \tag{52}$$

The needed consistent estimate of the deep parameters, $\hat{\pi} = [\hat{\beta}_x \quad \hat{\beta}_o' \quad \hat{\beta}_u]$, can now be obtained by the MLE based on this Poisson version of the DGP in (45).

3.3.2. The Conway-Maxwell Poisson (CMP) Model

For CMP model, the relevant form of (40) is

$$h_{(Y_{X^*}^*, S \mid W)}(Y_{X^*}^*,\, S,\, W,\, \pi,\, \delta) = \left(\int_{-W\delta}^{\infty} \mathcal{CMP}(Y_{X^*}^*;\, \lambda^*,\, \omega)\;\; g(X_u)\, dX_u \right)^{S} \times G(-W\delta)^{1-S} \tag{53}$$

where

$$\mathcal{CMP}(Y_{X^*}^*;\; \lambda^*,\, \omega) = \frac{(\lambda^*)^{Y_{X^*}^*}}{(Y_{X^*}^*\,!)^{\exp(\omega)}\, Z(\lambda^*, \exp(\omega))} \tag{54}$$

$$\lambda^* = \exp(X^*\beta_x + X_o\beta_o + X_u\beta_u)$$

$$Z(\lambda^*, \omega) = \sum_{j=0}^{\infty} \frac{\lambda^{*j}}{(j!)^{\exp(\omega)}}$$

The conditional mean of $(Y^*_{X^*} \mid X_o, X_u)$ for CMP model is

$$E[Y^*_{X^*} \mid X_o, X_u] = m(X^*,\ X_o,\ X_u, \pi) = \lambda^* \left(\sum_{j=1}^{\infty} \frac{j(\lambda^*)^{j-1}}{(j!)^{\exp(\omega)} Z(\lambda^*, \exp(\omega))} \right) \tag{55}$$

and the parameter vector is $\pi = [\beta \quad \omega]$ with $\beta' = [\beta_x \quad \beta_o' \quad \beta_u]$ and $-\infty < \omega < \infty$

Using the LIE and combining (55) and (37), the relevant form of (43) for the CMP case is given as

$$\begin{aligned} AIE_2(\Delta) = E_{X^{pre}, X_o, \Delta}\Bigg[\int_{-\infty}^{\infty} \Big\{ & m(X^{pre} + \Delta,\ X_o,\ X_u;\ \pi) \\ & - m(X^{pre},\ X_o,\ X_u;\ \pi) \Big\} g(X_u \mid W)\ dX_u \Bigg] \end{aligned} \tag{56}$$

Given a consistent estimate of π (say $\hat{\pi}$), the AIE can be consistently estimated using the following sample analog to (56)

$$\begin{aligned} \widehat{AIE_2(\Delta)} = \sum_{i=1}^{n} \frac{1}{n} \Big[& m(X_i^{pre} + \Delta_i,\ X_{oi},\ X_u;\ \hat{\pi}) \\ & - m(X_i^{pre},\ X_{oi},\ X_u;\ \hat{\pi}) \Big\} g(X_u \mid W)\ dX_u \Big] \end{aligned} \tag{57}$$

Turning now to consistent estimation of π, if the CIND condition holds, we can write the relevant DGP as in (45) as

$$h_{(Y^*_{X^*}, S \mid W)}(Y^*,\ S,\ W,\ \pi,\ \delta) = \left(\int_{-W\delta}^{\infty} \mathcal{CMP}(Y^*;\ \lambda,\ \omega)\ \ g(X_u)\ dX_u \right)^{S} \times G(-W\delta)^{1-S} \tag{58}$$

The consistent estimate of the deep parameters, $\hat{\pi}$, can now be obtained by the MLE based on (58).

3.4. Comparison of Alternative Sample Selection Conditional Potential Outcomes Models (CPOMs) using Simulated Data

In this section, I compare the count data models that accommodate under-dispersion and sample selection using simulated data. Similar to the approach in Chapter 2, the models are compared based on empirical bias and efficiency criteria. I follow the following procedure for the data simulation and estimation. First, I develop and test count sample selection maximum likelihood estimators (MLE) for the model parameters $\pi = [\beta, \omega]$ based on (46) for each of the CPOM models (poisson, CMM, RGP, and HP). Second, I develop estimators for the AIE of 1-year increment in year of education on DFS. Third, I write a Stata-Mata code for calculating the "true" AIE values for each data generation model design, which we discuss below. Finally, I place all of the data simulators and estimators in a replication context (with a "do" loop). The resultant do file will facilitate easy comparisons of the estimates (Poisson, COM, RGP, and HP) with respect to parameter estimation accuracy and efficiency. Performances of alternative models are compared taking the Absolute Proportional Bias (APB) and Average Absolute Proportional Bias (AAPB) defined in (35) and (36), respectively.

3.4.1. Sampling design

The sampling design varied along five values of the dispersion parameter for each model. i.e. the dispersion parameter ω is specified such that the generated data have different levels of dispersion. i use five different values of ω for different models because dispersion is defined differently in all the models. Simulation of the covariates $X = [X \;\; X_o \;\; X_u \;\; 1]$, is designed such that X and X_o are uniformly distributed with both mean 1 and variance of 3. Instrumental variables with mean and variance value of 1 also are generated. i also choose the true regression parameters $\beta' = [\beta_x \;\; \beta_o \;\; \beta_u \;\; c] = [0.25 \;\; -0.25 \;\; 0.25 \;\; 2.5]$, where $c = 0.25$ denotes the true intercept. I simulate data with sample size = 10,000 and number of repetitions = 100.

3.4.2. Estimation Results

I present the results with models with no selection (models that ignore selection) followed by CPOMs models that account for sample selection and compare the two types of models. Consistent estimation of the deep parameters in both cases was estimated according to the sampling design detailed above.

Table 7 presents the AIEs from non-sample selection models with exogeneous increment (Δ = 1) for increasing level of under-dispersion using data generated from CMP model. True AIE from CMP model is compared with estimated AIE (Δ) values from are the poisson, CMP and HP models. Comparing the poisson AIE estimates with the CMP and HP, I find that the AIE estimates from the sample DCR models are less biased, revealing the cost of ignoring under-dispersion. Table 8 presents sample selection models using data generated from CMP model. It is interesting to see that while the poisson AIE estimates are considerably different from each other sample selection and no sample selection cases, the AIE estimates from sample selection DCR models and no sample selection DCR models are not substantially different. I also find that the magnitude of the absolute proportional bias increases as under-dispersion increases.

The main take-away from the simulation results in Tables 7-10 is that the AIE estimates obtained from the sample selection DCR are substantially different from those obtained using models that ignore sample selection and do not account for under-dispersion/overdispersion. I take this as preliminary evidence that taking account of the sample selection and under/overdispersion may substantially affect the AIE estimation using real data. The following subsection is devoted for illustrating this.

3.5. Comparison of Alternative Sample Selection Conditional Potential Outcomes Models (CPOMs) using Real Data

This section illustrates the sample selection regression framework developed in the chapter by presenting an empirical analysis of the effect of a 1-year increment on wife's education on desired family size (DFS). The data for this analysis comes from the 3*2016 Ethiopia Demographic and Health Survey (EDHS). The data includes information collected from 35,081 women between 15 to 49 years of age. The data about the desired number of children were based on responses to the survey question: "*If you could go back to the time you did not have any children and could choose exactly the number of children to have in your whole life, how many would that be?*" Although a response to this question should be an integer value, nearly 14 percent gave the "At God's Will" (AGW) response. The remaining 86 percent gave numeric responses.

In the illustrative example, the $\mathbf{Y}$ is the number of children a woman desires to have in her life time and the $\mathbf{X}$ is the number of years of schooling of the mother. The variables included in X_o are: respondent's age; area of residence; marital status; religion; employment status; a dummy variable that indicates if the woman used a contraceptive method; and spouse's years of schooling. The descriptive statistics for these variables are presented in Table 13. The table includes descriptive statistics for the full sample and the two sub-samples (women who gave numeric responses, henceforth *the* $S=1$ *subsample* and women who gave the AGW response, henceforth *the* $S=0$ *sample*). Looking at the observable characteristics of the two subsamples in the descriptive statistics, I find that women in the $S=0$ subsample are older, less educated and have more children compared to those in the $S=1$ subsample. In addition to the observable characteristics, there could also be unobservable factors that could make the excluded women systematically different form the selected women. If unobservable factors determining the women's decision to give numeric responses are correlated with the unobservables influencing the outcome variable, selection issue arises. In the context of the illustrative example, the selection issue arises because the outcome is not observable for all members of the relevant population and sampling preclusion is not random. i.e. selection is based a woman's response to the desired fertility survey question. In this example, the selection variable, S, is a binary indicator defined as:

$$S = \begin{cases} 0 \text{ if a woman gives the AGW response} \\ 1 \text{ otherwise (a woman gives numeric response)} \end{cases}$$

To deal with the sample selection problem, in addition to the explanatory variables, I use two *identifying instrumental variables:* a dummy variable that takes a value 1 if the respondent ever took alcoholic drink and 0, otherwise and a dummy variable that takes value 1 if the woman knows her ovulatory cycle and 0 otherwise. These variables are chosen because they are highly correlated with the probability of giving the AGW response for the desired fertility question, but likely to affect the desired number of children.

3.5.1 Empirical Results

Here I present and discuss the MLE estimation results for the deep parameters and the AIE estimations from respective models. I first present the estimate for δ from selection equation obtained with a probit model applied to the full sample using the Stata 'probit' command.[12] This model represents the likelihood of giving numeric responses to the survey question on DFS. The positive coefficient for the Education variable indicates that educated women are more likely to numerically express their desired family size.

I then estimate both β' first with a poisson model and then under the CMP specification. For the poisson model, estimation of β' was conducted using the Stata 'poisson' command. Since there is no built-in Stata command for the CMP model,

[12] The full table of deep parameter estimates for this model are given in (Table 13).

estimation of the deep parameters β' and ω was done by manually coding the CMP regression model in Stata-Mata and using the 'm-optimize' command.[13]

After obtaining the estimates for the deep parameters, the AIE of a 1-year counterfactual increase in education was estimated using (51) and (57) for the poisson and CMP models, respectively. The formulations in (51) and (57), along with their correct asymptotic standard errors were coded in Mata [see Terza 2016 (a-c) and Terza 2017, 2018]. For the purpose of comparison, I also calculated the AIE from the conventional linear regression model [via ordinary least squares (OLS)], and Heckman twostep sample selection model. In the case of the linear models the relevant AIE is the OLS regression coefficient of the Education variable. The AIE estimates from different models are summarized in Table 12.

The AIE of a 1-year counterfactual increase in wife's education from the CMP Selection Model is -0.377 and it is statistically significant at the 1 percent significance level (see Table 12, Column 1). This estimate says that for a counterfactual 1-yr increase in education for the wife, family size reduces by 0.377 on average. For Ethiopia, a country with an estimated 20 million families, this effect translates to a 7,540,000 decline in the number of desired population size for an additional year of schooling for all wives in the country.

The AIE from the poisson Selection Model is -0.196 and is statistically significant. I find a 0.181 divergence from the AIE result for the CMP Selection Model. Translating this divergence to the nationwide level reveals that the divergence is

[13] The MLE results for the deep parameters from poisson and CMP models are given in Table 14.

substantial than it appears to be. Indeed, vis-a-vis the CMP Selection estimate, the Poisson Selection estimate understates the education effect by 3,620,000 people.

I next fit the Heckman twostep selection model. Unlike the CMP and poisson models, this model doesn't account for the fact that the outcome variable is count-valued. I find that a counterfactually increased 1 year of wife's education results in decline in family size by 0.078 (see Table 12 Column 2). The AIE estimate from this model diverges from that of the preferred model (the CMP Selection Model) by 0.299 (a 78.77 % difference). This difference in the AIE estimates appears to be substantial, leading to a large difference in a nation-wide effect. As can be seen in Table 12 (labeled 'Nationwide') the Heckman selection model underestimates the decline in population by 5,980,000.

As the empirical results show the estimates for AIE from non-selection methods are different from the selection models indicating that ignoring the selection could lead to biased estimates of the policy effect.

3.6. Summary and Conclusion

In this chapter, we present a FP endogenous sample selection modeling framework for count data models accommodating for all kinds of data dispersion. By incorporating the PO framework, we also provide a means of clearly and coherently defining causally interpretable PE parameter. Results based on the simulation study and an illustrative exercise show that accommodating the endogenous sample selection would lead to more precise AIE estimates. We also find that models that account for under-dispersion perform better in bias and efficiency measures compared to the poisson model, which neglects any kind of dispersion in the data. This result is consistent with the analysis in Chapter 2.

Chapter 4. Switching Regressions with Count Outcomes: Specification, Estimation and Causal Inference

In this chapter, I develop a regression-based PO approach to policy relevant causal inference to accommodate the possibility that some structural aspects of the model may vary with the value of a binary *switching variable*. Similar to the previous chapters, the focus is for the case in which the outcome variable of interest is count-valued. As an illustration, I once again consider estimation of the average incremental effect (AIE) of education on actual family size (AFS) using the same data used in Chapter 3. In the context of the illustrative example, the PO framework I propose in this chapter takes account of possible binary variation in regression structure that corresponds to differences in underlying attitudes toward fertility decision making that are reflected by responses to this survey question (AGW or not). As before, the models in this chapter take explicit account of the fact that the outcome of interest is count-valued and are designed to loosen the equi-dispersion constraint by implementing dispersion-flexible count-valued regression (DCR) specifications.

The organization of this chapter is very similar to the previous chapters. In the next section, I cast the switching regressions (SR) count outcome model in the PO framework and detail the formulation of the EP of interest (the AIE of education on AFS) in that framework. I then give the conditions under which the EP, which is based on a relevant counterfactual, can be consistently estimated with observed (factual) data using consistent estimates of the SR parameters (the *deep parameters* of the model). I finally apply the models and the methods to the estimation of the causal effect of a 1-year increment to the education levels of all wives in the relevant population on AFS family size. I find that education has a negative and significant effect on actual family size. Via relevant likelihood ratio tests, I reject the equi-dispersion and no-structural-switching null hypotheses. For the former, I reject in favor of under-dispersion. To assess the overall importance of jointly accounting for discreteness, under-dispersion and structural switching I estimated the model using OLS based the conventional linear regression specification. The OLS-based estimate of the AIE is negative and 22% smaller (in absolute value) than that the AIE estimate obtained from my proposed method. For Ethiopia, a country with an estimated 20 million households, this difference translates to a 680,000 shortfall in the expected population decrease vis-à-vis my estimate.

4.1. Switching Regressions with Count Outcomes from the Potential Outcomes Perspective

Terza (2019b) extends the GPOF in Terza (2019) to cases in which the structural aspects of the relevant conditional moments of both the counterfactual and factual versions of the $\mathbf{Y}$ correspond with a binary *switching variable* ($\mathbf{S}$) that is itself affected by the $\mathbf{X}$. Terza (2019b) also details EP specification, estimation and related inference that highlights the use of regression models and methods as means of reconciling the inherent disconnect between the counterfactual estimation objective and the factual data with which it is to be estimated. To develop the PO framework for the switching case I treat $\mathbf{Y}$ and the $\mathbf{S}$ as joint outcomes. Definition of two distinct versions of the pair $(\mathbf{Y}, \mathbf{S})$ is given as:

$(Y, S) \equiv$ the pair of random variables, whose first element is count-valued and whose second element is binary, representing the factual version of the joint distribution of the $\mathbf{Y}$ and the $\mathbf{S}$. [(The sampled values of the count-valued outcome and the switching variable are drawn from the distribution of (Y, S).]

and

$(Y_{X^*}, S_{X^*}) \equiv$ the pair of random variables representing the joint distribution of *potential outcomes*, defined as the distribution of values of the $(\mathbf{Y}, \mathbf{S})$ that would have manifested for a particular X^* (an exogenously mandated version of the $\mathbf{X}$).

The distinction between the factual and counterfactual versions of the $\mathbf{X}$ is as given in the Chapter 2.

For illustration, in this chapter, I seek to estimate the change in actual family size (the $\mathbf{Y}$) that can be attributed to an additional year of the wife's education ($\mathbf{X}$). Formally, I seek to estimate the *average incremental effect* (AIE) defined in (1). As discussed before, the EP in (1) is not directly estimable from data because pre X^{pre} and $Y_{X^{pre}}$ are counterfactual. Following Terza (2019 and 2019b), again, this gap between the estimation objective can be bridged via parametric specification of the conditional probability distribution of the relevant potential outcomes (Y_{X^*}, S_{X^*}) given a vector of control covariates (X_o). To this end, following the approach proposed by Terza

(2019b), I posit the following fully parametric (FP) structural specification for $[(Y_{X^*}, S_{X^*}) \mid X_o]$

$$\mathrm{pmf}((Y_{X^*}, S_{X^*}) \mid X_o) = f_{(Y_{X^*}, S_{X^*} \mid X_o)}(Y_{X^*}, S_{X^*}, X^*, X_o; \alpha, \kappa_1, \kappa_0)$$

$$= [1 - G(-\mathcal{X}^*\alpha)]\, f_{1\,(Y_{X^*} \mid X_o)}(Y_{X^*}, X^*, X_o, \kappa_1)^{S_{X^*}}$$

$$\times\, G(-\mathcal{X}^*\alpha)\, f_{0\,(Y_{X^*} \mid X_o)}(Y_{X^*}, X^*, X_o, \kappa_0)^{1-S_{X^*}} \qquad (59)$$

where

$\mathrm{pmf}((Y_{X^*}, S_{X^*}) \mid X_o) \equiv$ the joint conditional probability mass function (pmf) of (Y_{X^*}, S_{X^*}) given X_o

$[1 - G(-\mathcal{X}^*\alpha)] \equiv$ probability that $S_{X^*} = 1$, $G(\,)$ being a known cumulative distribution function (cdf)

$f_{1\,(Y_{X^*} \mid X_o)}(Y_{X^*}, X^*, X_o, \kappa_1) \equiv$ is the relevant form of the count pmf of Y_{X^*} when $S_{X^*} = 1$

$f_{0\,(Y_{X^*} \mid X_o)}(Y_{X^*}, X^*, X_o, \kappa_0) \equiv$ is the relevant form of the count pmf of Y_{X^*} when $S_{X^*} = 0$

$\mathcal{X}^* = [X^* \quad X_o]$

and the "deep" parameters of the model are $\alpha = [\alpha_x \;\; \alpha_o']$, κ_1 and κ_0. Alternative specifications for $f_{1\,(Y_{X^*}\,|\,X_o)}(\;)$, $f_{0\,(Y_{X^*}\,|\,X_o)}(\;)$ and G( ) will be discussed in the next section in the context of the education/family size illustration. From (59) it follows that[14]

$$
\begin{aligned}
m(X^*, X_o, \pi) &= E[Y_{X^*} \,|\, X_o] \\
&= [1 - G(-(X^*\alpha_x + X_o\alpha_+))]m_1(X^*, X_o, \kappa_1) \\
&\qquad + G(-(X^*\alpha_x + V\alpha_+))m_0(X^*, X_o, \kappa_0) \qquad (60)
\end{aligned}
$$

where

$$m_1(X^*, X_o, \kappa_1) = E_1[Y_{X^*} \,|\, X_o] = \int_{-\infty}^{\infty} Y_{X^*} \; f_{1\,(Y_{X^*}\,|\,X_o)}(Y_{X^*}, X^*, X_o, \kappa_1)\; dY_{X^*}$$

$$m_0(X^*, X_o, \kappa_0) = E_0[Y_{X^*} \,|\, X_o] = \int_{-\infty}^{\infty} Y_{X^*} \; f_{0\,(Y_{X^*}\,|\,X_o)}(Y_{X^*}, X^*, X_o, \kappa_0)\; dY_{X^*}$$

and $\pi = [\alpha' \;\; \kappa_1' \;\; \kappa_0']$. Using (60) and the law of iterated expectation (LIE), the equivalent of (1) for this chapter can be rewritten as

$$AIE_3(\Delta) = E[m(X^{pre} + \Delta, X_o; \pi)] - E[m(X^{pre}, X_o; \pi)]. \qquad (61)$$

[14]See Terza (2019b) for the derivation of the conditional mean function (60) from the joint pmf of $[(Y_{X^*}, S_{X^*}) \,|\, X_o]$ in (59).

where $\pi = [\alpha' \quad \kappa_1' \quad \kappa_0']$. It is relatively easy to show that under very general conditions, given a consistent estimate of π (say $\hat{\pi}$) the AIE can be consistently estimated using the following sample analog to (61)

$$\widehat{\mathrm{AIE}(\Delta)} = \sum_{i=1}^{n} \frac{1}{n} \left\{ m(X_i^{pre} + \Delta_i, X_{oi}; \hat{\pi}) - m(X_i^{pre}, X_{oi}; \hat{\pi}) \right\} \tag{62}$$

where X_i^{pre} and Δ_i are the exogenously determined values of X^{pre} and Δ for the i^{th} observation in a sample of size n (i=1,..., n); and X_{oi} is the sampled value of X_o. As is made clear by (62), consistent estimation of the EP (61) hinges on the existence of a consistent estimate of π. With this in mind and noting the conditional independence of (Y_{X^*}, S_{X^*}) and X given X_o (the CIND assumption), the following is legitimate

$$\begin{aligned} \mathrm{pdf}(Y,S \mid X_o) &= f_{(Y_{X^*}, S_{X^*} \mid X_o)}(Y, S, X, X_o; \alpha, \tau_1, \tau_0) \\ &= [1 - G(-\mathcal{X}\alpha)] f_{1\,(Y_{X^*} \mid X_o)}(Y, X, X_o, \kappa_1)^S \\ &\quad \times G(-\mathcal{X}\alpha) f_{0\,(Y_{X^*} \mid X_o)}(Y, X, X_o, \kappa_0)^{1-S} \end{aligned} \tag{63}$$

where $\mathcal{X} = [X \quad X_o]$. In other words, under the CIND condition, the relevant DGP can be obtained from (59) by replacing the counterfactual random variables Y_{X^*} S_{X^*} and X^* with the observable random variables Y, S and X, respectively. From (63) it follows that $\pi = [\alpha' \quad \kappa_1' \quad \kappa_0']$ can be consistently estimated as the MLE obtained as

$$\hat{\pi} = \underset{\breve{\pi}=[\breve{\alpha}' \;\; \breve{\kappa}_1' \;\; \breve{\kappa}_0']}{\arg\max} \sum_{i=1}^{n} q(\breve{\pi}, Z_i) \qquad (64)$$

where

$$q(\breve{\pi}, Z_i) = \ln[f_{(Y_{X^*}, S_{X^*} \mid X_o)}(Y_i, S_i, X_i, X_{oi}; \breve{\alpha}, \breve{\kappa}_1, \breve{\kappa}_0)] \qquad (65)$$

and $Z_i = [Y_i \quad S_i \quad X_i \quad X_{oi}]$. It is useful to note that

$$\begin{aligned} q(\breve{\pi}, Z_i) &= \ln[f_{(Y_{X^*}, S_{X^*} \mid X_o)}(Y_i, S_i, X_i, X_{oi}; \breve{\alpha}, \breve{\kappa}_1, \breve{\kappa}_0) \\ &= S_i \ln[1 - G(-X_i\breve{\alpha})] + S_i \ln\left[f_{1\,(Y_{X^*} \mid X_o)}(Y_i, X_i, X_{oi}, \breve{\kappa}_1)\right] \\ &\qquad + (1-S_i) \ln[G(-X_i\breve{\alpha})] + (1-S_i) \ln\left[f_{0\,(Y_{X^*} \mid X_o)}(Y_i, X_i, X_{oi}, \breve{\kappa}_0)\right] \\ &= q_1(\breve{\alpha}, S_i, X_i, X_{oi}) + q_2(\breve{\kappa}_1, Y_i, X_i, X_{oi}) + q_3(\breve{\kappa}_0, Y_i, X_i, X_{oi}) \qquad (66) \end{aligned}$$

where

$$q_1(\breve{\alpha}, S_i, X_i, X_{oi}) = S_i \ln[1 - G(-X_i\breve{\alpha})] + (1-S_i) \ln[G(-X_i\breve{\alpha})]$$

$$q_2(\breve{\kappa}_1, Y_i, X_i, X_{oi}) = S_i \ln\left[f_{1\,(Y_{X^*} \mid X_o)}(Y_i, X_i, X_{oi}, \breve{\kappa}_1)\right]$$

$$q_3(\breve{\kappa}_0, Y_i, X_i, X_{oi}) = (1-S_i) \ln\left[f_{0\,(Y_{X^*} \mid X_o)}(Y_i, X_i, X_{oi}, \breve{\kappa}_0)\right].$$

So (64) can be rewritten as

$$\begin{aligned} \hat{\pi} &= \underset{\breve{\pi}=[\breve{\alpha}' \;\; \breve{\kappa}_1' \;\; \breve{\kappa}_0']}{\arg\max} \sum_{i=1}^{n} q(\breve{\pi}, Z_i) \\ &= \underset{\breve{\alpha}}{\arg\max} \sum_{i=1}^{n_1} q_1(\breve{\alpha}, S_i, X_i, X_{oi}) + \underset{\breve{\kappa}_1}{\arg\max} \sum_{i=1}^{n_1} q_2(\breve{\kappa}_1, Y_i, X_i, X_{oi}) \end{aligned}$$

$$+\arg\max_{\breve{\kappa}_0} \sum_{i=1}^{n_0} q_3(\breve{\kappa}_0, Y_i, X_i, X_{oi})$$

(67)

where n_1 and n_0 denote the sizes of the subsamples for which $S_i = 1$ and $S_i = 0$. Therefore, α, κ_1, and κ_0 can be separately estimated: the MLE for α by the appropriate binary response model applied to the full sample; and the MLEs for κ_1, and κ_0 by the appropriate count-valued outcome models applied to the respective subsamples for which $S_i = 1$ and $S_i = 0$.

4.2. Alternative Count Outcome Conditional Potential Outcomes Models (CPOMs) with switching

In equation (59) I give the generic SR count-outcomes PO specification. Particular versions of this PO specification correspond with alternative forms for G( ), $f_{1\,(Y_{X^*} \mid X_o)}(\ )$ and $f_{0\,(Y_{X^*} \mid X_o)}(\ )$. I maintain the assumption that G ( ) ≡ Φ( ) [probit model for switching] and consider two versions of the f functions: a) the conventional poisson specification; and b) a more flexible generalized version of the Poison – the Conway-Maxwell Poisson.

4.2.1. The Poisson Model

For the poisson model

$$f_{1\,(Y_{X^*}\,|\,X_o)}(Y_{X^*},\,X^*,\,X_o,\,\kappa_1) = \mathrm{POI}(Y;\,\lambda_1^*) \tag{68}$$

where

$$\mathrm{POI}(Y;\;\lambda_1^*) = \frac{(\lambda_1^*)^{Y_{X^*}}\exp(-\lambda_1^*)}{Y_{X^*}!}$$

$$E_1[Y_{X^*}\;|\;X_o] = \lambda_1^* = \exp(X^*\beta_{x1} + X_o\beta_{o1})$$

and the parameter vector is $\kappa_1 = \beta_1$ with $\beta_1' = [\beta_{x1}\quad \beta_{o1}']$. The forms of $f_{0\,(Y_{X^*}\,|\,X_o)}(\;\;)$, λ_0^*, $\kappa_0 = \beta_0$ and $\beta_0' = [\beta_{x0}\quad \beta_{o0}']$ are analogously specified. Using (62), and these specifications I obtain

$$E[Y_{X*}\;|\;X_o] = \Phi(X^*\alpha_x + X_o\alpha_o)\lambda_1^* + [1-\Phi(X^*\alpha_x + X_o\alpha_o)]\lambda_0^*. \tag{69}$$

Using the LIE and combining (69) and (1) the relevant form of (61) for the poisson case can be given as

$$\mathrm{AIE}(\Delta) = E\left[\mu(X^{pre}, X_o,\;\Delta) - \mu(X^{pre}, X_o,\;0)\right] \tag{70}$$

where

$$\mu(X^{pre}, X_o, \Delta) = \Phi^{pre}(\Delta)\lambda_1^{pre}(\Delta) + [1-\Phi^{pre}(\Delta)]\lambda_0^{pre}(\Delta)$$

$$\lambda_1^{pre}(\Delta) = \exp([X^{pre} + \Delta]\beta_{x1} + X_o\beta_{o1})$$

$$\lambda_0^{pre}(\Delta) = \exp([X^{pre} + \Delta]\beta_{x0} + X_o\beta_{o0})$$

$$\Phi^{pre}(\Delta) = \Phi([X^{pre} + \Delta]\alpha_x + X_o\alpha_o).$$

and the full vector of deep parameters of the model is $\pi = [\alpha' \quad \kappa_1' \quad \kappa_0']$ with $\alpha' = [\alpha_x \quad \alpha_o']$. As it is clear from earlier discussions, given consistent estimates of the deep parameters (say $\hat{\pi} = [\hat{\alpha}' \quad \hat{\kappa}_1' \quad \hat{\kappa}_0']$) , (61) can be consistently estimated using the following sample analog

$$\widehat{\mathrm{AIE}(\Delta)} = \sum_{i=1}^{n} \frac{1}{n}\left\{\hat{\mu}(X_i^{pre}, X_{oi}, \Delta_i) - \hat{\mu}(X_i^{pre}, X_{oi}, 0)\right\} \tag{71}$$

where

$$\hat{\mu}(X_i^{pre}, X_{oi}, \Delta_i) = \hat{\Phi}_i^{pre}(\Delta_i)\hat{\lambda}_{1i}^{pre}(\Delta_i) + [1 - \hat{\Phi}_i^{pre}(\Delta_i)]\hat{\lambda}_{0i}^{pre}(\Delta_i)$$

$$\hat{\lambda}_{1i}^{pre}(\Delta_i) = \exp([X_i^{pre} + \Delta_i]\hat{\beta}_{x1} + X_{oi}\hat{\beta}_{o1})$$

$$\hat{\lambda}_{0i}^{pre}(\Delta_i) = \exp([X_i^{pre} + \Delta_i]\hat{\beta}_{x0} + X_{oi}\hat{\beta}_{o0})$$

and

$$\hat{\Phi}^{pre}(\Delta_i) = \Phi([X_i^{pre} + \Delta_i]\hat{\alpha}_x + X_{oi}\hat{\alpha}_o).$$

Turning now to consistent estimation of π, if the CIND condition holds, the relevant DGP as in (63) with $G(-X\alpha)$ replaced by $1 - \Phi(X\alpha_x + X_o\alpha_o)$ can be written as

$$f_{1\,(Y_{X^*} \mid X_o)}(Y, X, X_o, \kappa_1) = \mathrm{POI}(Y; \lambda_1) \tag{72}$$

and

$$f_{0\,(Y_{X^*} \mid X_o)}(Y, X, X_o, \kappa_0) = \mathrm{POI}(Y; \lambda_0)$$

where

$$E_1[Y \mid X_o] = \lambda_1 = \exp(X\beta_{x1} + X_o\beta_{o1})$$

and

$$E_0[Y \mid X_o] = \lambda_0 = \exp(X\beta_{x0} + X_o\beta_{o0}).$$

The needed consistent estimate of the deep parameters, $\hat{\pi} = [\hat{\alpha}' \quad \hat{\kappa}_1' \quad \hat{\kappa}_0']$, can now be obtained by the MLE based on this poisson version of the DGP in (63).

4.2.2. The Conway-Maxwell Poisson (CMP) Model

For CMP model

$$f_{1\,(Y_{X^*} \mid X_o)}(Y_{X^*}, X^*, X_o, \kappa_1) = \mathcal{CMP}(Y; \lambda_1^*, \omega_1) \tag{73}$$

where

$$\mathcal{CMP}(Y;\ \lambda_1^*, \omega_1) = \frac{(\lambda_1^*)^Y}{(Y!)^{\exp(\omega_1)} Z(\lambda_1^*, \exp(\omega_1))} \tag{74}$$

$$\lambda_1^* = \exp(X^*\beta_{x1} + X_o\beta_{o1})$$

$$Z(\lambda^*, \omega) = \sum_{j=0}^{\infty} \frac{\lambda^{*j}}{(j!)^{\exp(\omega)}}$$

$$E_1[Y_{X^*} \mid X_o] = = \lambda_1^* \left(\sum_{j=1}^{\infty} \frac{j(\lambda_1^*)^{j-1}}{(j!)^{\exp(\omega_1)} Z(\lambda_1^*, \exp(\omega_1))} \right) \tag{75}$$

and the parameter vector is $\kappa_1 = [\beta_1 \;\; \omega_1]$ with $\beta_1^{'} = [\beta_{x1} \;\; \beta_{o1}^{'}]$ and $-\infty < \omega < \infty$ .The forms of $f_{0\,(Y_{X^*}\,|\,X_o)}(\;\;)$, λ_0^*, $m_0(\lambda_0^*, \omega_0)$, $\kappa_0 = [\beta_0 \;\; \omega_0]$ and $\beta_0^{'} = [\beta_{x0} \;\; \beta_{o0}^{'}]$ are analogously specified. Using (60), and these specifications I obtain

$$E[Y_{X^*} \mid X_o] = \Phi(X^*\alpha_x + X_o\alpha_o)\, m_1(\lambda_1^*, \omega_1) + [1 - \Phi(X^*\alpha_x + X_o\alpha_o)]\, m_0(\lambda_0^*, \omega_1) \tag{76}$$

Using the LIE and combining (76) and (1) I obtain the relevant form of (61) for the CMP case as

$$\text{AIE}(\Delta) = E\left[m(X^{pre}, X_o, \Delta) - m(X^{pre}, X_o, 0) \right] \tag{77}$$

where

$$m(X^{pre}, X_o, \Delta) = \Phi^{pre}(\Delta)\, m_1(\lambda_1^{pre}(\Delta), \omega_1) + [1 - \Phi^{pre}(\Delta)]\, m_0(\lambda_0^{pre}(\Delta), \omega_0)$$

$$\lambda_1^{pre}(\Delta) = \exp([X^{pre} + \Delta]\beta_{x1} + X_o\beta_{o1})$$

$$\lambda_0^{pre}(\Delta) = \exp([X^{pre} + \Delta]\beta_{x0} + X_o\beta_{o0})$$

$$\Phi^{pre}(\Delta) = \Phi([X^{pre} + \Delta]\alpha_x + X_o\alpha_o)$$

and the full vector of deep parameters of the model is $\pi = [\alpha' \;\; \kappa_1' \;\; \kappa_0' \;\; \omega_1 \;\; \omega_0]$ with $\alpha' = [\alpha_x \;\; \alpha_o']$. Given consistent estimates of the deep parameters (say $\hat{\pi} = [\hat{\alpha}' \;\; \hat{\kappa}_1' \;\; \hat{\kappa}_0' \;\; \hat{\omega}_1 \;\; \hat{\omega}_0]$), (61) can be consistently estimated using the following sample analog

$$\widehat{\text{AIE}(\Delta)} = \sum_{i=1}^{n} \frac{1}{n} \left\{ \hat{m}(X_i^{pre}, X_{oi}, \Delta_i) - \hat{m}(X_i^{pre}, X_{oi}, 0) \right\} \tag{78}$$

where

$$\hat{m}(X_i^{pre}, X_{oi}, \Delta_i) = \hat{\Phi}_i^{pre}(\Delta_i) m_1(\hat{\lambda}_{1i}^{pre}(\Delta_i), \hat{\omega}_1) + [1 - \hat{\Phi}_i^{pre}(\Delta_i)] m_{0i}(\hat{\lambda}_{0i}^{pre}(\Delta_i), \hat{\omega}_0)$$

$$\hat{\lambda}_{1i}^{pre}(\Delta_i) = \exp([X_i^{pre} + \Delta_i]\hat{\beta}_{x1} + X_{oi}\hat{\beta}_{o1})$$

$$\hat{\lambda}_{0i}^{pre}(\Delta_i) = \exp([X_i^{pre} + \Delta_i]\hat{\beta}_{x0} + X_{oi}\hat{\beta}_{o0})$$

and

$$\hat{\Phi}^{pre}(\Delta_i) = \Phi([X_i^{pre} + \Delta_i]\hat{\alpha}_x + X_{oi}\hat{\alpha}_o).$$

Turning now to consistent estimation of π, if the CIND condition holds, relevant DGP as in (63), with $G(-X\alpha)$ replaced by $1 - \Phi(X\alpha_x + X_o\alpha_o)$, can be written as

$$f_{1\,(Y_{X^*} \mid X_o)}(Y_{X^*}, X^*, X_o, \kappa_1) = \mathcal{CMP}(Y; \lambda_1, \omega_1) \tag{79}$$

and

$$f_{0\,(Y_{X^*} \mid X_o)}(Y_{X^*}, X^*, X_o, \kappa_0) = \mathcal{CMP}(Y; \lambda_0, \omega_0) \tag{80}$$

where

$$\mathcal{CMP}(Y;\ \lambda_1^*, \omega_1) = \frac{(\lambda_1)^Y}{(Y!)^{\exp(\omega_1)} Z(\lambda_1, \exp(\omega_1))}$$

$$\mathcal{CMP}(Y;\ \lambda_0^*, \omega_0) = \frac{(\lambda_0)^Y}{(Y!)^{\exp(\omega_0)} Z(\lambda_0, \exp(\omega_0))}.$$

The consistent estimate of the deep parameters, $\hat{\pi} = [\hat{\alpha}' \quad \hat{\kappa}_1' \quad \hat{\kappa}_0' \quad \hat{\omega}_1 \quad \hat{\omega}_0]$, can now be obtained by the MLE based on this CMP version of the DGP in (63).

Another nice property of the CMP is that it nests the standard poisson distribution when the dispersion parameter $\omega=0$. The data is over dispersed if $\omega < 0$, and under-dispersed if $\omega > 0$. The fact that CMP reduces to the poisson when $\omega = 0$ allows for a simple statistical test of under- or -over dispersion.

4.3. Comparison of Alternative Switching CPOMs Using Real Data

This section illustrates the switching regression framework developed in section 4.3 by presenting an empirical analysis of the effect of the wife's education on AFS. I once again use the 2016 EDHS.[15] The objective in this chapter is to estimate the AIE of 1-year increment on wife's education on AFS (not DFS). The switching modeling aspect arises in this case from the fact that the data on the AFS comes from population of women with possible systematic differences in their preferences towards their family sizes as reflected in their response to the DFS survey question: "*If you could go back to the time you did not have any children and could choose exactly the number of children to have in your whole life, how many would that be?*". To the best of my knowledge, there has been no research dealing with the issue in SR framework. The model and method developed in this chapter are designed to accommodate this modeling aspect of fertility data in addition to accounting for the under-dispersion and for the fact that $\mathbf{Y}$ is a count valued outcome

In the context of the current illustrative example the $\mathbf{Y}$ is the number of children in a household and the $\mathbf{X}$ is the number of years of schooling of the mother. The variables included in X_o are: respondent's age; area of residence; marital status; religion; employment status; a dummy variable that indicates if the woman used a contraceptive method; and spouse's years of schooling. The descriptive statistics for these variables are

[15] See Chapter 3 for detailed description of the data.

presented in Table 1. The table includes descriptive statistics for the full sample and the two sub-samples (women who gave numeric responses, henceforth *the* $S=1$ *subsample* and women who gave the AGW response, henceforth *the* $S=0$ *sample*). Comparing the two subsamples based on the descriptive statistics, I find that women in the $S=0$ subsample are older, less educated and have more children compared to those in the $S=1$ subsample.

4.3.1. Estimation Results

Here I present and discuss the MLE estimation results for the deep parameters and the AIE estimations from respective models. I first note that the log-likelihood function in (67), on which the MLE estimation is based, is additively separable. This allows separate MLE estimations of the deep parameters $\pi = [\alpha' \quad \kappa_1' \quad \kappa_0']$. First, I estimate α' with a probit model applied to the full sample using the Stata 'probit' command.[16] This model represents the likelihood of giving numeric responses to the survey question on DFS. The positive coefficient for the Education variable indicates that educated women are more likely to numerically express their desired family size.

I then estimate both κ_1 and κ_0 separately first with a poisson model and then under the CMP specification. For the poisson model, estimation of κ_1 and κ_0 was conducted using the Stata 'poisson' command applied to the $S=1$ and $S=0$ subsamples separately. Since there is no built-in Stata command for the CMP model, estimation of

[16] The full table of deep parameter estimates for this model are given in the Appendix (Table 3).

the deep parameters κ_1, κ_2, ω_1 and ω_0 was done by manually coding the CMP regression model in Stata-Mata and using the 'm-optimize' command.[17]

After obtaining the estimates for the deep parameters, the AIE of a 1-year counterfactual increase in education was estimated using (71) and (78) for the poisson and CMP models, respectively. The formulations in (71) and (78), along with their correct asymptotic standard errors were coded in Mata [see Terza 2016 (a-c) and Terza 2017, 2018]. For the purpose of comparison, I also calculated the AIE from the conventional linear regression model [via ordinary least squares (OLS)], in which case the relevant AIE is the OLS regression coefficient of the Education variable. The AIE estimates from different models are summarized in Table 15.

The AIE of a 1-year counterfactual increase in wife's education from the CMP Switching Model is -0.155 and it is statistically significant at the 1 percent significance level (see Table 15, Column 1).[18] This estimate says that for a counterfactual 1-yr increase in education for the wife, family size reduces by 0.155 on average. For Ethiopia, a country with an estimated 20 million families, this effect translates to a 3,100,00 decline in population for an additional year of schooling for all wives in the country. From the CMP Switching Model I also obtain positive and statistically significant coefficient estimates for the dispersion parameters $(\omega_1, \omega_0 > 0)$. As discussed earlier, for this model, a positive value of the dispersion parameter corresponds to under-dispersion in the data. Since the poisson model is nested in CMP when the disperson parameter is 0,

---

[17] The MLE results for the deep parameters from poisson and CMP models are given in the Appendix (Table 14).

[18] The full table of deep parameter estimates for this model are given in the (Table 14).

the positive and significant values for $\hat{\omega}_1$ and $\hat{\omega}_0$ implicitly reject the Poisson Switching Model in favor of the CMP Switching Model.

I next fit the conventional linear model that ignores all of the modeling aspects of the SR model. I find that a counterfactually increased 1 year of wife's education results in decline in family size by 0.121 (see Table 15 Column 2).[19] The AIE estimate from this model diverges from that of my preferred model (the CMP switching model) by 0.034 (a 21.94 difference). Although this difference in the AIE estimates appears to be nominally small, it corresponds to a large downward divergence when translated to a nation-wide effect. As can be seen in the fifth row of Table 15 (labeled 'Nationwide') the linear model underestimates the decline in population by 680,000.

The AIE from the Poisson Switching Model is -0.160 and is statistically significant.[20] Here again, I get an apparently small 0.005 nominal divergence from the AIE result for the CMP Switching Model (a 3.23 difference). Translating this divergence to the nationwide level, however, reveals that it is not trivial. Indeed, vis-a-vis the CMP Switching estimate, the Poisson Switching estimate overstates the education effect by 100,000 people. As discussed earlier, the CMP Switching results for $\hat{\omega}_1$ and $\hat{\omega}_0$ indicate rejection of the Poisson Switching Model. As a means of validating this result, I conducted a conventional likelihood ratio (LR) test of the null hypothesis that the Poisson Switching Model is the true model. Such a test is available to us, of course, because the null model (H_o :Poisson Switching) is nested in the alternative hypothesis (H_A: CMP Switching). As can be seen in the last row (third column) of Table15, the value of the LR

[19] The full table of deep parameter estimates for this model are given in the Table 14.
[20] The full table of deep parameter estimates for this model are given in the Table 14

statistic is 3567.088 which is significantly different from 0 at any reasonable size of the test. Thus, the Poisson Switching Model is rejected in favor of the CMP Switching Model based on this likelihood ratio test.

Finally, I estimate the AIE using the deep parameter results from a version of the CMP model that does not involve switching. I obtain the CMP No Switching version of the model by setting $\alpha = 0$ and $\kappa_0 = \kappa_1 = \kappa$ in (59). Imposing these restrictions to the CMP Switching Model yields the following version of (59)

$$\begin{aligned} f_{1\,(Y_{X^*} \mid X_o)}(Y_{X^*}, X^*, X_o, \kappa_1) &= f_{0\,(Y_{X^*} \mid X_o)}(Y_{X^*}, X^*, X_o, \kappa_0) \\ &= f_{(Y_{X^*} \mid X_o)}(Y_{X^*}, X^*, X_o, \kappa) = \mathcal{CMP}(Y_{X^*}; \lambda^*, \omega) \end{aligned} \tag{81}$$

where

$$\mathcal{CMP}(Y_{X^*};\ \lambda^*, \omega) = \frac{(\lambda^*)^{Y_{X^*}}}{(Y\,!)^{\exp(\omega)}\, Z(\lambda^*, \exp(\omega))}$$

$$\lambda^* = \exp(X^* \beta_x + X_o \beta_o)$$

$$Z(\lambda^*, \omega) = \sum_{j=0}^{\infty} \frac{\lambda^{*j}}{(j!)^{\exp(\omega)}}$$

$$E[Y_{X^*} \mid X_o] = = \lambda^* \left(\sum_{j=1}^{\infty} \frac{j(\lambda^*)^{j-1}}{(j!)^{\exp(\omega)}\, Z(\lambda^*, \exp(\omega))} \right). \tag{82}$$

Using the LIE and combining (83) and (1) the relevant form of (63) for the CMP case can be given as

$$AIE(\Delta) = E\left[m(X^{pre}, X_o,\ \Delta) - m(X^{pre}, X_o,\ 0) \right] \tag{83}$$

where

$$m(X^{pre}, X_o, \Delta) = m(\lambda^{pre}(\Delta), \omega)$$

$$\lambda^{pre}(\Delta) = \exp([X^{pre} + \Delta]\beta_x + X_o\beta_o)$$

and the full vector of deep parameters of the model is $\pi = [\kappa' \;\; \omega]$. Given consistent estimates of the deep parameters (say $\hat{\pi} = [\hat{\kappa}' \;\; \hat{\omega}]$) , (83) can be consistently estimated using the following sample analog

$$\widehat{\text{AIE}(\Delta)} = \sum_{i=1}^{n} \frac{1}{n} \left\{ \hat{m}(X_i^{pre}, X_{oi}, \Delta_i) - \hat{m}(X_i^{pre}, X_{oi}, 0) \right\} \tag{84}$$

where

$$\hat{m}(X_i^{pre}, X_{oi}, \Delta_i) = m(\hat{\lambda}_i^{pre}(\Delta_i), \hat{\omega})$$

$$\hat{\lambda}_i^{pre}(\Delta_i) = \exp([X_i^{pre} + \Delta_i]\hat{\beta}_x + X_{oi}\hat{\beta}_o).$$

Turning now to consistent estimation of π, I again note that if the CIND condition holds (see Terza, 2019, 2019b), the relevant DGP for the CMP No Switching Model can be written as

$$f_{(Y_{X^*} \mid X_o)}(Y, X, X_o, \kappa) = \mathcal{CMP}(Y; \lambda, \omega) \tag{85}$$

where

$$\mathcal{CMP}(Y;\; \lambda, \omega) = \frac{(\lambda)^{Y}}{(Y!)^{\exp(\omega)} Z(\lambda, \exp(\omega))}.$$

A consistent estimate of the deep parameter vector ($\hat{\pi} = [\hat{\kappa}' \;\; \hat{\omega}]$), can now be obtained as the MLE based on (85).

Using this CMP, No Switching Model I estimate that the 1-year counterfactual increment in education would lead to a decline in family size by 0.166 (first row last column of Table 15).[21] This estimate diverges from the CMP Switching Model estimate by 0.011 for an average family. When projected to the entire nation, this divergence would mean that the CMP No Switching Model overestimates the decline in population size by 220,000 (second row, last column of Table 15 – a difference that may be substantial from a fertility policy perspective. The likelihood ratio test statistic for the CMP no switching null (LR= 215.702) confirms this conclusion by rejecting the CMP No Switching model in favor of the CMP Switching Model. This simple test in this case is possible because the CMP No Switching Model is nested in the CMP Switching Model, as shown above.

4.4. Summary, Discussion and Conclusion

A new regression-based approach is developed for the specification and estimation of the effect of a presumed causal variable on a count-valued outcome of interest. The underlying regression is cast in the PO framework to ensure causal interpretability of the EP. Conditions required for estimating the EP with observed (factual) data using the appropriately specified DGP are discussed. The MLE estimation technique for consistently estimating the model parameters is detailed. An empirical study that analyzes the effect of a policy mandated 1-year increment in the wife's education on AFS is used to illustrate the new approach. The regression-based model presented in this chapter, in addition to accounting for the fact that the outcome of

[21]Full table of deep parameter estimates for this model are given in the Appendix (see Table 4)

interest is a count valued, also accommodates possible structural aspects of the model that may vary with the value of a binary *switching variable.* In the illustrative empirical analysis, the sample selection model accounts for possible differences in fertility decisions between women who gave numeric responses and those who gave the AGW response for the survey question on desired family size. Moreover, my approach loosens the equi-dispersion constraint, imposed by the conventional poisson model, using the more flexible CMP model, which allows for any of the three types of dispersion in the data for the count outcome (equi-, over-, and under-dispersion).

Applying my regression-based causal effect estimator to DHS data for Ethiopia, I find that a counterfactually imposed 1-year increment to the education levels of all wives in the relevant population would cause a .15 decrease in family size. I have shown that this estimate, although it might appear to be small, could translate to a large nation-wide decrease in population. Likelihood ratio tests reject the equi-dispersion and no-structural-switching null hypotheses suggesting the need to account for structural switching and the under-dispersion existed in the data. Moreover, my estimate of the education effect substantively differs from the OLS estimate, as would follow from a conventional linear regression specification. In particular, I find that OLS underestimates the effect of an additional year of schooling on family size by 0.034, making the nation-wide effect of an estimated 680,000 decrease in population size.

Policy makers in developing countries seeking to design effective policies for influencing population growth should find this analysis useful. It should be noted, however, that my regression-based PO specification relies on the assumption that the policy variable of interest ($\mathbf{X}$) is exogenous. In practice, there may be several unobservable confounders that could invalidate this assumption making causal interpretability of my estimates questionable. As an extension to this work I plan to incorporate such endogeneity into our regression framework.

Chapter 5: Summary, Discussion and Conclusions

Count data regression models are applied to cases in which the outcome of interest is nonnegative integer. Conventional count data regression models are mainly based on standard poisson which is of limited relevance for some applications due to its restrictive equidispersion assumption. In this dissertation, I develop a regression-based models for the specification and estimation of the effect of a presumed causal variable on a count-valued outcome of interest. The regression-based models presented in the dissertation are designed to loosen the equi-dispersion assumption of the poisson model. The models in this dissertation also account for sample selection bias in the context of count data and accommodate possible structural aspects of the model that may vary with the value of a binary *switching variable.* This work contributes to the literature in the following specific ways. First, specification, estimation and inference for such models are placed in a PO framework, thereby making explicit the requisite conditions for causal interpretability of the treatment effect parameters and estimates. Second, a simulation study of alternative count data models is conducted. The simulation study showed that accounting for under-dispersion in the data has clear potential gains in accuracy with the estimation of EP of interest. Separate tests for under-dispersion also indicate that the DCR models are more appropriate when under-dispersion exists in the data and leads to more efficient parameter estimates. Finally, an empirical study of household fertility choice is used to for the purpose of illustrating the aforementioned count data modeling aspects and the DCR models were applied to estimate AIE of 1-year increment in wife's education on AFS and DFS. This dissertation should be viewed as a first step toward the development and application of DCR for the estimation EPs in count data contexts. A number of possible extensions of this work are reserved for future research.

Chapter 6: Tables

Table 1: Average Incremental Effect (AIE) Estimates using CMP Data

Level of dispersion (ω)	True $AIE(\Delta_L)$	$\widetilde{AIE}(\Delta_L)$ Poisson	$\widetilde{AIE}(\Delta_L)$ RGP	$\widetilde{AIE}(\Delta_L)$ CMP	$\widetilde{AIE}(\Delta_L)$ HP	Absolute %bias (Poisson)	Absolute %bias (RGP)	Absolute %bias (CMP)	Absolute %bias (HP)
1.5	.233	.240	.076	.235	.240	2.9% (9.3%)	6.7%	.73%	3.0% (10.6%)
2.5	.125	.130	.035	.125	.128	4.1% (13.3%)	7.1%	.18%	2.7% (8.4%)
5	.068	.073	.029	.068	.068	6.4% (23.1%)	131.2%	.41%	.05% (1.3%)
7.5	.058	.062	--	.057	.057	7.8% (34.1)	--	.90%	.87% (.07%)
10	.056	.060	--	.055	.055	8.4% (37.1%)	--	.80%	.88% (.007%)

*Percentage in the brackets is the Average Absolute Proportional Bias as in (36)

Table 2: Average Incremental Effect (AIE) Estimates using RGP Data *

Level of dispersion (ω)	True $AIE(\Delta_L)$	$\widehat{AIE(\Delta_L)}$ Poisson	$\widehat{AIE(\Delta_L)}$ RGP	$\widehat{AIE(\Delta_L)}$ CMP	$\widehat{AIE(\Delta_L)}$ HP	Absolute %bias (Poisson)	Absolute %bias (RGP)	Absolute %bias (CMP)	Absolute %bias (HP)
-.05	.438	443	.443	.436	.447	1.2% (.3%)	1.25%	.4% (4.6%)	2.12% (5.6%)
-.1	.438	.443	.444	.430	.453	1.23% (2.3%)	1.3%	1.7% (1.1%)	103.5% (23.2%)
-.15	.438	.444	.445	.427	.461	1.35% (8.5%)	1.62%	2.5% (9.3%)	5.31% (62.3%)
-.25	.438	.335	.278	.324	.345	23.5% (35.9%)	36.5%	26.0% (29.0%)	21.6% (42.2%)
-.5	.438	-.046	-.035	-.046	-.046	110.5% --	108.2% --	110.6% --	110.6% --

*Percentage in the brackets is the Average Absolute Proportional Bias as in (36)

Table 3: Average Incremental Effect (AIE) Estimates using HP Data

Level of dispersion (ω)	True $AIE(\Delta_L)$	$\widetilde{AIE(\Delta_L)}$ Poisson	$\widetilde{AIE(\Delta_L)}$ RGP	$\widetilde{AIE(\Delta_L)}$ CMP	$\widetilde{AIE(\Delta_L)}$ HP	Absolute %bias (Poisson)	Absolute %bias (RGP)	Absolute %bias (CMP)	Absolute %bias (HP)
.75	.453	.457	.343	.454	.459	5.1%	.9%	6.2%	1.2%
.5	.462	.460	.389	.454	.467	4.4%	3.0%	5.6%	1.1%
.25	.460	.448	.422	.442	.465	2.5%	8.1%	3.8%	1.1%
.1	.450	.430	.436	.424	.455	4.4%	3.3%	5.6%	1.1%
.05	.444	.422	.440	.417	.450	5.1%	.9%	6.2%	1.2%

Table 4: Summary Statistics Wang and Famoye (1997) Data

Variables	Observation	Mean	Std. Dev.	Min	Max
Education	1954	13.34	2.22	5	19
Employed	1954	.72	.44	0	1
Age of wife	1954	31.17	5.27	18	40
Family income	1954	43.97	27.48	2.11	348.9
Race	1954	.732	.44	0	1
City	1954	.527	.49	0	1
Children	1954	1.69	1.17	0	9

Table 5: Coefficient Estimates using Wang and Famoye (1997) Data ***

Y	Poisson	RGP	CMP	HP
Education	-.027 (-3.22) [.001]	-.024 (-3.36) [.001]	-.037 (-3.73) [.000]	-.032 (-3.50) [.000]
Employed	-.322 (-8.55) [.000]	-.314 (-9.73) [.000]	-.436 (-9.46) [.000]	-.385 (-9.12) [.000]
Age of wife	.026 (7.68) [.000]	.024 (8.36) [.000]	.035 (8.58) [.000]	.031 (8.23) [.000]
Family income	-.002 (-3.31) [.001]	-.002 (-3.70) [.000]	-.003 (-3.81) [.000]	-.003 (-3.61) [.000]
Race	-.167 (-4.23) [.000]	-.165 (-4.86) [.000]	-.227 (-4.86) [.000]	-.2005 (-4.59) [.000]
City	-.023 (-0.64) [.520]	-.017 (-0.57) [.566]	-.031 (-.75) [.451]	-.027 (-0.70) [.482]
Constant	.520 (3.60) [.000]	.539 (4.30) [.000]	.962 (5.49) [.000]	.249 (1.54) [.123]
ω	--	-.074 (-1.80) [.000]	1.511	.536
$\upsilon = \ln(\omega)$	--	--	.413 (-1.80) [.000]	-.622 (-6.88) [.000]

n=1954

*** Values without bracket are coefficient estimates, ( )—t-stat, [ ]-P-value

Table 6: Average Incremental Effect (AIE) using Wang and Famoye (1997) Data

Model	AIE Estimates	Asymp. se	Asymp. t-stat	P-value
Poisson	-.045	.014	-3.25	.001
RGP	-.041	.012	-3.40	.0006
CMP	-.046	.012	-3.79	.0001
Hyper-P	-.045	.012	-3.55	.0003

n=1954

Table 7: Average Incremental Effect (AIE) Estimates using CMP Data—No Sample Selection Case

Level of dispersion (ω)	True $AIE(\Delta_L)$	$\widetilde{AIE(\Delta_L)}$ Poisson NO SS	$\widetilde{AIE(\Delta_L)}$ CMP NO SS	$\widetilde{AIE(\Delta_L)}$ HP NOSS	Absolute %bias (Poisson)	Absolute %bias (CMP)	Absolute %bias (HP)
1.5	.234	.237	.232	.237	1.28%	.84%	1.26%
2.5	.125	.128	.123	.127	2.40%	1.56%	1.57%
5	.069	.072	.067	.068	4.35%	2.78%	1.47%
7.5	.058	.062	.057	.057	6.90%	1.61%	1.75%
10	.056	.060	.055	.055	7.14%	1.67%	1.82%

Table 8: Average Incremental Effect (AIE) Estimates using CMP Data—Sample Selection Case

Level of dispersion (ω)	True $AIE(\Delta_L)$	$\widetilde{AIE(\Delta_L)}$ Poisson SS	$\widetilde{AIE(\Delta_L)}$ CMP SS	$\widetilde{AIE(\Delta_L)}$ HP SS	Absolute %bias (Poisson)	Absolute %bias (CMP)	Absolute %bias (HP)
1.5	.267	.384	.221	.365	43.8%	17.6%	36.7%
2.5	.137	.350	.110	--	155.4%	19.7%	--
5	.070	.336	.045	.253	380%	35.7%	261%
7.5	.058	.334	.034	.119	475%	41.4%	105%
10	.055	.333	.032	-.390	505%	41.8%	809%

Table 9: Average Incremental Effect (AIE) Estimates using HP Data—No Sample Selection Case

Level of dispersion (ω)	True $AIE(\Delta_L)$	$\widetilde{AIE(\Delta_L)}$ Poisson NO SS	$\widetilde{AIE(\Delta_L)}$ CMP NO SS	$\widetilde{AIE(\Delta_L)}$ HP NO SS	Absolute %bias (Poisson) NO SS	Absolute %bias (CMP) NO SS	Absolute %bias (HP) NO SS
.75	.454	.451	.448	.453	.66%	1.32%	.22%
.5	.463	.454	.448	.461	1.94%	3.24%	.43%
.25	.460	.444	.436	.459	3.50%	5.22%	.22%
.1	.450	.424	.418	.449	5.78%	7.11%	.22%
.05	.445	.416	.410	.444	6.52%	7.86%	.23%

Table 10: Average Incremental Effect (AIE) Estimates using HP Data—Sample Selection Case

Level of dispersion (ω)	True $AIE(\Delta_L)$	$\widetilde{AIE(\Delta_L)}$ Poisson SS	$\widetilde{AIE(\Delta_L)}$ CMP SS	$\widetilde{AIE(\Delta_L)}$ HP SS	Absolute %bias (Poisson) SS	Absolute %bias (CMP) SS	Absolute %bias (HP) SS
.75	.528	.512	.435	.468	3.03%	17.61%	11.36%
.5	.533	.608	.439	.468	14.07%	17.63%	12.20%
.25	.529	.768	.472	.458	45.18%	10.78%	13.42%
.1	.521	.902	.517	.456	73.12%	.77%	12.48%
.05	.518	.955	.532	.458	84.36%	2.70%	11.58%

Table 11: Descriptive Statistics EDHS Data

	Full Sample		Numeric ($S=1$)		AGW ($S=0$)	
	Mean	SD	Mean	SD	Mean	SD
Number of children	4.971	2.257	4.875	2.254	5.580	2.181
Education	1.571	3.331	1.712	3.471	0.678	2.037
Age	34.654	7.418	34.443	7.411	35.993	7.322
Rural	0.835	0.371	0.826	0.379	0.894	0.308
Muslim	0.491	0.499	0.469	0.499	0.630	0.483
Married	0.989	0.105	0.988	0.108	0.992	0.087
Employed	0.440	0.496	0.438	0.496	0.449	0.498
Contraceptives	0.743	0.437	0.726	0.446	0.853	0.354
Husband education	3.108	4.485	3.269	4.573	2.085	3.724
Obs.	$N = 35081$		$n_1 = 30307$		$n_0 = 4774$	

Table 12: Average Incremental Effect (AIE) of 1-yr Counterfactual Increase in Wife's Education on DFS

CMP Selection	Poison Selection	Linear Selection	CMP No Selection	Poisson No Selection	Linear No Selection
-0.377*** (0.030)	-0.196*** (0.014)	-0.078*** (0.011)	-0.080*** (0.011)	-0.076*** (0.005)	- 0.068*** (0.011)
Nationwide AIE					
7,540,000	3,920,000	1,560,000	1,600,000	1, 520,000	1,360,000
	Divergence from Estimated AIE for CMP Selection Model				
%	48.01%	79.31%	78.77%	79.84%	81.96%
Per Family	-0.181	-0.299	-0.297	-0.301	-0.309
Nationwide	3,620,000	5, 980,000	5,940,000	6,020,000	6,180,000

Table 13: Regression Results (Probit and OLS)

	Probit	OLS (No Switch)
	$\hat{\alpha}$	$\hat{\kappa}_1 = \hat{\kappa}_0 = \hat{\kappa}$
Education	0.045***	-0.121***
	(0.004)	(0.004)
Age	-0.013***	0.168***
	(0.001)	(0.001)
Rural	-0.094***	0.796***
	(0.029)	(0.029)
Muslim	-0.312***	0.536***
	(0.019)	(0.020)
Married	-0.244***	0.728***
	(0.092)	(0.087)
Employed	-0.147***	-0.112***
	(0.018)	(0.019)
Contraceptives	-0.259***	0.178***
	(0.023)	(0.022)
Spouse Education	0.005*	0.013***
	(0.003)	(0.003)
Constant	2.268***	-2.442***
	(0.109)	(0.105)
Obs.	N = 35081	N = 35081

Note: Standard error in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 14: Regression Results (Poisson and CMP models)

	Poisson No Switch	Poisson S=1	Poisson S = 0	CMP No Switch	CMP S=1	CMP S = 0
	$\hat{\kappa}_1 = \hat{\kappa}_0 = \hat{\kappa}$	$\hat{\kappa}_1$	$\hat{\kappa}_0$	$\hat{\kappa}_1 = \hat{\kappa}_0 = \hat{\kappa}$	$\hat{\kappa}_1$	$\hat{\kappa}_0$
Education	-0.035***	-0.034***	-0.023***	-0.054***	-0.054***	-0.036***
	(0.001)	(0.001)	(0.004)	(0.001)	(0.001)	(0.005)
Age	0.034***	0.035***	0.031***	0.055***	0.056***	0.049***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)
Rural	0.168***	0.181***	0.073***	0.269***	0.291***	0.119***
	(0.008)	(0.009)	(0.023)	(0.010)	(0.011)	(0.028)
Muslim	0.105***	0.109***	0.059***	0.169***	0.176***	0.095***
	(0.005)	(0.006)	(0.015)	(0.006)	(0.007)	(0.017)
Married	0.184***	0.200***	0.032	0.292***	0.319***	0.053
	(0.025)	(0.027)	(0.074)	(0.032)	(0.034)	(0.095)
Employed	-0.019***	-0.017***	-0.042***	-0.030***	-0.027***	-0.069***
	(0.005)	(0.005)	(0.014)	(0.006)	(0.007)	(0.017)
Contraceptives	0.039***	0.042***	-0.014	0.062***	0.068***	-0.022
	(0.006)	(0.006)	(0.018)	(0.008)	(0.008)	(0.023)
Spouse Education	0.004***	0.003***	0.008***	0.006***	0.004***	0.012***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.003)
Constant	0.027	-0.026	0.476***	0.231***	0.152***	0.942***
	(0.030)	(0.032)	(0.086)	(0.038)	(0.040)	(0.112)
Omega				0.516***	0.522***	0.520***
				(0.008)	(0.009)	(0.021)
Obs.	$N = 35081$	$n_1 = 30307$	$n_0 = 4774$	$N = 35081$	$n_1 = 30307$	$n_0 = 4774$

Table 15: Average Incremental Effect (AIE) of 1-yr Counterfactual Increase
in Wife’s Education on AFS

CMP Switching	Linear No Switching	Poisson Switching	CMP No Switching
-0.155*** (0.007) $\hat{\omega}_1 = 0.522$*** (0.009) $\hat{\omega}_0 = 0.520$*** (0.021)	-0.121*** (0.004)	-0.160*** (0.006)	-0.166*** (0.004)
Nationwide AIE			
3,100,000	2,420,000	3,200,000	3, 320,000
	Divergence from Estimated AIE for CMP Switching Model		
%	21.94	3.23	7.10
Per Family	-0.034	.005	.011
Nationwide	-680,000	100,000	220,000
		Likelihood Ratio Statistics $\mathbf{H_A}$: CMP Switching	
		$\mathbf{H_0}$:Poisson Switching	$\mathbf{H_0}$:CMP No Switching
		LR= 3567.088 ***	LR= 215.702 ***

Note: standard error in parentheses; *** p<0.01

References

Cameron, A. C., & Trivedi, P. K. (2005): "Microeconometrics: methods and applications," Cambridge university press.

Conway, R. W., & Maxwell, W. L. (1962): "A queuing model with state dependent service rates," *Journal of Industrial Engineering*, *12*(2), 132-136.

Cygan-Rehm, K (2011): "Between here and there. Immigrant fertility patterns in Germany," *University of Erlangen-Nuremberg*

Famoye, F. (1993): "Restricted generalized Poisson regression model," *Communications in Statistics-Theory and Methods*, *22*(5), 1335-1354.

Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995): "Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models," *Psychological bulletin*, *118*(3), 392.

Greene, W. (2001): "FIML estimation of sample selection models for count data," In *Economic Theory, Dynamics and Markets* (pp. 73-91). Springer, Boston, MA.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimation for such models. *Annals of Economic and Social Measurement*, 5 (4), 475-492.

Islam et al (2013): "Predictors of the number of under-five malnourished children in Bangladesh: application of the generalized Poisson regression model," *BMC Public Health* http://www.biomedcentral.com/1471-2458/13/11

Jensen, E. (1985): "Desired fertility, the "up to god" response, and sample selection bias," *Demography*, *22*(3), 445-454.

Khazraee, S.H. (2014): "Application of the Hyper-Poisson Generalized Linear Model for Analyzing Motor Vehicle Crashes," *Risk Analysis*, DOI: 10.1111/risa.12296

Mayer, J., & Riphahn, R. T. (2000): "Fertility assimilation of immigrants: Evidence from count data models," *Journal of population Economics*, *13*(2), 241-261.

Riley, A. P., Hermalin, A. I., & Rosero-Bixby, L. (1993): "A new look at the determinants of nonnumeric response to desired family size: the case of Costa Rica," *Demography*, *30*(2), 159-174.

Sáez-Castillo, A. J., & Conde-Sánchez, A. (2013): " A hyper-Poisson regression model for overdispersed and underdispersed count data," *Computational Statistics & Data Analysis*, *61*, 148-157.

Terza, J.V. (2019): "Regression-Based Causal Analysis from the Potential Outcomes Perspective," *Journal of Econometric Methods*, published online ahead of print, DOI**:** https://doi.org/10.1515/jem-2018-0030.

Terza, J.V. (2019b): "Causal Inference and Switching Regressions," Unpublished Manuscript, Department of Economics, Indiana University Purdue University Indianapolis.

Terza, J.V. (2017): "Causal Effect Estimation and Inference Using Stata," the *Stata Journal*, 939-961.

Terza, J.V. (2018): "Even Simpler Standard Errors for Two-Stage Optimization Estimators: Mata Implementation via the DERIV Command," Presented at the Stata Conference, Columbus, OH, July.
https://www.stata.com/meeting/columbus18/slides/columbus18_Terza.pdf

Terza, J.V. 1998, "Estimating count data models with endogenous switching: sample selection and endogenous treatment effects," *Journal of Econometrics* 84, 129-154.

Terza, J.V. (2009): "Parametric Nonlinear Regression with Endogenous Switching," *Econometric Reviews*, 28, 555-580.

Terza, J.V. (2016a): "Simpler Standard Errors for Two-Stage Optimization Estimators," the *Stata Journal*, 16, 368-385.

________(2016b): "Inference Using Sample Means of Parametric Nonlinear Data Transformations," *Health Services Research*, 51, 1109-1113.

________(2016c): "Supplementary Appendix to 'Inference Using Sample Means of Parametric Nonlinear Data Transformations,'" *Health Services Research*, DOI: 10.1111/1475-6773.12494.

Wang, W. and Famoye, Felix (1997): "Modeling Household Fertility Decisions with Generalized Poisson Regression," *Journal of Population Economics*, 10, 273-28.

Winkelmann, R. (2008): "Econometric analysis of count data*,"* Springer Science & Business Media.

Winkelmann, R., & Zimmermann, K. F. (1992): "Robust poisson regression," In *Advances in GLIM and Statistical Modelling* (pp. 201-206). Springer, New York, NY.

Winkelmann, R., & Zimmermann, K. F. (1994): “Count data models for demographic data,” *Mathematical Population Studies*, *4*(3), 205-221.

Winkelmann and Zimmermann (2009): “Count data models for demographic data,” Mathematical Population Studies: *An International Journal of Mathematical Demography,4:3,205-221* http://www.tandfonline.com/action/showCitFormats?doi=10.1080/08898489409525374

Wooldridge, J. M. (2010): “Econometric analysis of cross section and panel data,” MIT press.

Curriculum Vitae

Wondimu Samuel Manalew

Education

- Ph.D. in Economics, Indiana University, Indianapolis, 2020
- MA in Economics, Vanderbilt University, Nashville, 2014
- MSc. in Economics, Addis Ababa University, Ethiopia, 2011
- BA in Economics, Haramaya University, Ethiopia, 2006

Research Fields

- Health Economics, Applied Econometrics, Development Economics

Publications and Working Papers

- Mekonen, S., Manalew, W.S, & Ambelu, A. (2014): "Importance of labelling and patient knowledge to ensure proper care during drug dispensing," A case study from a tertiary hospital in Ethiopia. *Open Journal of Preventive Medicine*, *4*(01), 1.
- Manalew W.S & Tennekoon V.S (2019). "Dirty Hands on Troubled Waters: Sanitation, Access to Water and Child Health in Ethiopia" forthcoming, *Review of Development Economics*
- "Private Aid, Disasters and Development: Evidence from OECD and US Data Sources" (with Una Osili) [*Under Review*]
- "Does Education Reduce Child Marriage? Evidence from a Natural Experiment", *Working Paper*

Work Experience

- Statistical Analyst, Indiana University, Lilli Family School of Philanthropy, 2016
- Graduate Research Assistant, Energy Balance Laboratory, Diabetes Center at Vanderbilt University Medical Center, Nashville, TN (2012-2014)

Teaching Experience

- Teaching Fellow, AEA Summer Program at Michigan State University, Summer 2018
- Instructor, Center for Talented Youth, Johns Hopkins University, Summer 2017
- Instructor, Indiana University (IUPUI), 2016 to 2019
- Lecturer, Haramaya University, Ethiopia, 2006 to 2011

Awards and Scholarships

- Carlin Award for Outstanding Graduate Student Paper in Empirical Economics, IUPUI, April 2018
- Graduate Student Travel Fellowship Award, IUPUI, November 2018
- Annual Health Econometrics Workshop Student Scholarship, September 2018
- Loretta Lunsford Scholarship, IUPUI, 2018/2019 Academic Year
- Graduate and Professional Educational Grant, IUPUI, January 2018
- GPED Award for Academic Excellence, Vanderbilt University, May 2014
- GPED Special Award for Service to the Graduate Program, Vanderbilt University, May 2014
- Vanderbilt University, Full Tuition Scholarship, 2012-2014